

METHODS FOR USING DOUBLE-MUTANT RNA POLYMERASES WITH REDUCED DISCRIMINATION BETWEEN NON-CANONICAL AND CANONICAL NUCLEOSIDE TRIPHOSPHATES

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application is a continuation-in-part of U.S. Patent Application Serial No 10/404,726, filed April 1, 2003 which claims priority from U.S. Serial No. 09/965,346, filed September 27, 2001, which is a continuation of 09/643,189 filed August 21, 2000, which is a continuation of U.S. Serial No. 09/100,803, (U.S. Patent No. 6,107,037) filed June 19, 1998, which is a divisional of U.S. Serial No. 08/713,331 (U.S. Patent No. 5,849,546), filed September 13, 1996. This application also claims priority from U.S. provisional patent application Serial No. 60/420,009 filed October 21, 2002.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with United States government support awarded by the following agencies: NIH GM52522. The United States has certain rights in this invention.

FIELD OF THE INVENTION

[0003] The field of the present invention is methods for producing nucleic acid molecules containing at least one non-canonical nucleotide and for characterizing nucleic acid molecules by synthesizing nucleic acid molecules containing at least one non-canonical nucleotide *in vitro* using double-mutant nucleic acid polymerases having at least a 10-fold reduced discrimination between non-canonical ribonucleoside triphosphates and canonical ribonucleoside-5'-triphosphates as substrates compared to the corresponding wild-type enzymes.

BACKGROUND OF THE INVENTION

[0004] There are a number of procedures commonly used in the art for *in vitro* synthesis of nucleic acid molecules, including both DNA and RNA. For example, one may use an *in vitro* transcription reaction to synthesize RNA from a DNA template present in the reaction. T7-type RNA polymerases, such as T7 RNA polymerase, T3 RNA polymerase or SP6 RNA polymerase, are commonly used in such reactions, although many other RNA polymerases may also be used. Usually, but not always, synthesis of RNA is *de novo* (i.e., unprimed), and usually, but not always, transcription is initiated at a sequence in the template that is specifically recognized by

the RNA polymerase, termed a "promoter" or a "promoter sequence". A method for *in vitro* transcription is presented herein.

[0005] Procedures for *in vitro* nucleic acid synthesis are also commonly used in the art to amplify nucleic acid molecules, including both DNA and RNA. For example, transcriptions using RNA polymerases are an integral part of "nucleic acid sequence-based amplification" (NASBA), "self-sustained sequence replication" (3SR), and "transcription-mediated amplification" (TMA) Hill, C.S., 1996, three similar methods for amplifying nucleic acid molecules *in vitro*.

[0006] By way of example, all or a specific portion of an RNA molecule may be amplified using NASBA (Compton, et al., 1991) or 3SR (Fahy, et al., 1991) by isothermal incubation of a sample RNA in a buffer containing two primers (a first primer complementary to the RNA molecule and encoding a promoter sequence for an RNA polymerase and a second primer complementary to the 3'-end of the first cDNA strand resulting from reverse transcription of the RNA molecule), an RNA- and DNA-dependent DNA polymerase which also has RNase H activity (or a separate RNase H enzyme), all four canonical 2'-deoxynucleoside-5'-triphosphates (dATP, dCTP, dGTP and dTTP), an RNA polymerase that recognizes the promoter sequence of the first primer, and all four canonical ribonucleoside-5'-triphosphates (rATP, rCTP, rGTP and rUTP).

[0007] A first cDNA strand is synthesized by extension of the first primer by reverse transcription. Then, the RNase H digests the RNA of the resulting DNA:RNA hybrid, and the second primer primes synthesis of the second cDNA strand. The RNA polymerase then transcribes the resultant double-stranded DNA (ds-DNA) molecule from the RNA polymerase promoter sequence, making many more copies of RNA, which in turn, are reversed transcribed into cDNA and the process begins all over again. This series of reactions, from ds-DNA through RNA intermediates to more ds-DNA, continues in a self-sustained way until reaction components are exhausted or the enzymes are inactivated. DNA samples can also be amplified by other variations of NASBA or 3SR or TMA.

[0008] Another nucleic acid amplification method involving DNA synthesis is the polymerase chain reaction (PCR).

[0009] By way of example, a specific portion of a DNA molecule may be amplified using PCR by temperature cycling of a sample DNA in a buffer containing two primers (one primer complementary to each of the DNA strands and which, together, flank the DNA sequence of interest), a thermostable DNA polymerase, and all four canonical 2'-deoxynucleoside-5'-triphosphates (dATP, dCTP, dGTP and dTTP). The specific nucleic acid sequence is

geometrically amplified during each of about 30 cycles of denaturation (e.g., at 95°C), annealing of the two primers (e.g., at 55°C), and extension of the primers by the DNA polymerase (e.g., at 70°C), so that up to about a billion copies of the nucleic acid sequence are obtained. RNA may be similarly amplified using one of several protocols for (reverse transcription-PCR) RT-PCR, such as, for example, by carrying out the reaction using a thermostable DNA polymerase which also has reverse transcriptase activity (Myers and Gelfand, 1991).

[0010] The polymerase chain reaction, discussed above, is the subject of numerous publications and patents, including, for example: Mullis, K.B., et al., U.S. Pat. No. 4,683,202 and U.S. Pat. No. 4,965,188.

[0011] A variety of procedures for using *in vitro* nucleic acid synthesis for characterizing nucleic acid molecules, including both DNA and RNA, also are known in the art.

[0012] There are many reasons for characterizing nucleic acid molecules. For example, genes are rapidly being identified and characterized which are causative or related to many human, animal and plant diseases. Even within any particular gene, numerous mutations are being identified that are responsible for particular pathological conditions. Thus, although many methods for detection of both known and unknown mutations have been developed (e.g., see Cotton, 1993), our growing knowledge of human and other genomes makes it increasingly important to develop new, better, and faster methods for characterizing nucleic acids. Besides diagnostic uses, improved methods for rapidly characterizing nucleic acids will also be useful in many other areas, including human forensics, paternity testing, animal and plant breeding, tissue typing, screening for smuggling of endangered species, and biological research.

[0013] One of the most informative ways to characterize a DNA molecule is to determine its nucleotide sequence. The most commonly used method for sequencing DNA at this time (Sanger, et al., 1977) uses a DNA polymerase to produce differently sized fragments depending on the positions (sequence) of the four bases (A = Adenine; C = Cytidine; G = Guanine; and T = Thymine) within the DNA to be sequenced. In this method, the DNA to be sequenced is used as a template for *in vitro* DNA synthesis. RNA may also be used as a template if a polymerase with RNA-directed DNA polymerase (i.e., reverse transcriptase) activity is used. In addition to all four of the deoxynucleotides (dATP, dCTP, dGTP and dTTP), a 2',3'-dideoxynucleotide is also included in each *in vitro* DNA synthesis reaction at a concentration that will result in random substitution of a small percentage of a normal nucleotide by the corresponding dideoxynucleotide. Thus, each DNA synthesis reaction yields a mixture of DNA fragments of different lengths corresponding to chain termination wherever the dideoxynucleotide was incorporated in place of the normal deoxynucleotide.

[0014] The DNA fragments are labeled, either radioactively or non-radioactively, by one of several methods known in the art and the label(s) may be incorporated into the DNA by extension of a labeled primer, or by incorporation of a labeled deoxy- or dideoxy- nucleotide. By carrying out DNA synthesis reactions for each of the four dideoxynucleotides (ddATP, ddCTP, ddGTP or ddTTP), then separating the products of each reaction in adjacent lanes of a denaturing polyacrylamide gel or in the same lane of a gel if different distinguishable labels are used for each reaction, and detecting those products by one of several methods, the sequence of the DNA template can be read directly. Radioactively-labeled products may be read by analyzing an exposed X-ray film. Alternatively, other methods commonly known in the art for detecting DNA molecules labeled with fluorescent, chemiluminescent or other non-radioactive moieties may be used.

[0015] Because 2',3'-dideoxynucleotides (ddNTPs), including even ddNTPs with modified nucleic acid bases, can be used as substrates for many DNA polymerases, Sanger's dideoxy- sequencing method is widely used. Recently, Tabor and Richardson (EP application 942034331, 1994) reported that mutations at specific sites in many DNA polymerases improved the ability of these mutant enzymes to accept ddNTPs as substrates, thereby leading to improved DNA polymerases for DNA sequencing using the Sanger method.

[0016] Nucleic acid sequencing provides the highest degree of certainty as to the identity of a particular nucleic acid. Also, nucleic acid sequencing permits one to detect mutations in a gene even if the site of the mutation is unknown. Sequencing data may even provide enough information to permit an estimation of the clinical significance of a particular mutation or of a variation in the sequence.

[0017] Cycle sequencing is a variation of Sanger sequencing that achieves a linear amplification of the sequencing signal by using a thermostable DNA polymerase and repeating chain terminating DNA synthesis during each of multiple rounds of denaturation of a template DNA (e.g., at 95°C), annealing of a single primer oligonucleotide (e.g., at 55°C), and extension of the primer (e.g., at 70°C).

[0018] Other methods for sequencing nucleic acids are also known besides the Sanger method. For example, Barnes described a method for sequencing DNA by partial ribo-substitution (Barnes, W.M., 1977). In this method, a pre-labeled primer was extended *in vitro* on a template DNA to be sequenced in each of four reactions containing a wild-type DNA polymerase in the presence of Mn^{2+} , all four canonical 2'-deoxyribonucleoside triphosphates, and one of four ribonucleoside triphosphates under deoxy-/ribo- nucleotide ratios and conditions that result in about 2% ribonucleotide substitution at each position. After alkali treatment to cleave

the synthetic DNA at the positions of partial ribosubstitution, the sequence was determined by analyzing the fragments resulting from each reaction following electrophoresis on a denaturing polyacrylamide gel.

[0019] Although most methods for sequencing nucleic acids employ DNA polymerases, some work has also been reported on the use of T7 RNAP and SP6 RNAP for transcription sequencing of DNA templates beginning at the respective T7 or SP6 promoter sequence using 3'-deoxyribonucleoside-5'-triphosphates (Axelrod, V.D., and Kramer, F.R., 1985), and Q-Beta replicase for sequencing single-stranded RNA templates (Kramer, F.R., and Mills, D.R., 1978). Also, 3'-O-methyl-ribonucleoside-5'-triphosphates have been used for sequencing DNA templates with *E. coli* RNA polymerase ((Axelrod, V.D., et al., 1978). None of these techniques is commonly used at present, perhaps in part, due to the difficulty to obtain the 3'-deoxy- and 3'-O-methyl- nucleoside triphosphate substrates, while 2',3'-dideoxy-ribonucleoside-5'-triphosphates that are commercially available have not been found to be substrates for wild-type (W.T.) RNA polymerases.

[0020] In view of the numerous applications involving *in vitro* nucleic acid synthesis known in the art, it is useful to consider the properties of the key nucleic acid polymerase reagents which make these procedures possible, and which, if modified in their essential properties, might improve these procedures.

[0021] One classification of nucleic acid polymerases relies on their different template specificities (RNA or DNA), substrate specificities (rNTPs or dNTPs), and mode of initiation (*de novo* or primed). These designations usually refer to the template and substrate specificities displayed *in vivo* during the fulfillment of a polymerase's biological function.

[0022] *In vitro*, polymerases can display novel activities, albeit with reduced efficiency and/or under non-physiological conditions. *E. coli* DNA-directed DNA polymerase I, for example, can use RNA as a template, although there is a concomitant ~100-fold increase in dNTP K_m (Ricchetti and Buc, 1993). T7 DNA-directed RNA polymerase can also use RNA as a template (Konarska and Sharp, 1989). These are not exceptional observations because it is a general property of polymerases that they display relaxed template specificity, at least *in vitro*.

[0023] While template specificity may be relaxed, polymerase substrate specificity is normally extremely stringent. T7 DNAP, for example, displays at least 2,000-fold selectivity for dNTPs over rNTPs, even in Mn^{2+} buffer which relaxes the ability of the polymerase to discriminate between dNTPs and ddNTPs (Tabor and Richardson, 1989).

[0024] It has been reported that transcripts synthesized by a T7 RNAP Y639F mutant *in vivo* yielded 1/2-1/3 of the protein per transcript compared to transcripts synthesized by the wild-

type enzyme (Makarova, et al., 1995). The latter phenotype was unique to the Y639F mutant amongst a number of other active site mutants examined for *in vivo* expression, and indicated that Y639F transcripts contained a defect that led to their being inefficiently translated.

[0025] A polymerase with an altered substrate specificity would be useful in many molecular biological applications, such as creating a nucleic acid molecule comprising at least one non-canonical nucleotide.

SUMMARY OF THE INVENTION

[0026] The present invention is summarized as methods and kits for synthesizing a nucleic acid molecule comprising at least one non-canonical nucleoside triphosphate using a double-mutant polymerase having a reduced discrimination between canonical and non-canonical substrates is disclosed. The method comprises incubating a template nucleic acid in a reaction mixture comprising the mutant nucleic acid polymerase and the appropriate canonical and non-canonical nucleoside triphosphates which are desired substrates for the mutant nucleic acid polymerase. The present invention is also a method of determining the sequence of a nucleic acid molecule using the mutant polymerase to create a nucleic acid molecule comprising at least one non-canonical nucleoside triphosphate. As such disclosed herein is the identification of double-mutant polymerases, such as T7-type RNAPs, that display the ability to use NTPs with various non-canonical groups or modifications at the ribose 2' or other positions. The double mutations occur in tyrosine 639 within motif B (Delarue, et al., 1990) of T7 RNAP and histidine 784 of T7 RNAP. The Examples below describe the location of the mutations in other T7-type RNAPs.

[0027] The present invention requires a double-mutant polymerase with a reduced discrimination between canonical and non-canonical nucleoside triphosphates.

[0028] In one embodiment, the present invention is a method for synthesizing a nucleic acid molecule that comprises at least one non-canonical nucleotide. This method comprises the steps of incubating a template nucleic acid in a reaction mixture suitable for nucleic acid polymerization containing the double-mutant nucleic acid polymerase and the appropriate canonical and non-canonical nucleoside triphosphates which are substrates for a mutant nucleic acid polymerase and which are desired to be incorporated into the synthesized nucleic acid molecule.

[0029] In an especially suitable form of this method, the synthesized nucleic acid molecule has an altered susceptibility to a nuclease compared to a nucleic acid which could be

synthesized using the corresponding non-mutant nucleic acid polymerase with canonical nucleoside triphosphates.

[0030] The present invention is also a method for determining the sequence of a nucleic acid molecule using a double-mutant RNA polymerase.

[0031] The method comprises synthesizing a nucleic acid molecule, either *de novo* from a promoter, or by extending a primer annealed to the template molecule in four separate reactions. The four separate reactions each have all 4 rNTPs and a portion of a ddNTP, or have all 4 dNTPs and a portion of a ddNTP, or have 4 2'-fluorine-substituted NTPs and a portion of a ddNTP. Chain termination will occur and the products may be evaluated so that the sequence of the template molecule may be deduced. In one embodiment of this method, the reactions which include a ddNTP occur in the same reaction mixture and are linked to a method for nucleic acid amplification, including, but not limited to, NASBA, 3SR, TMA, or other similar methods.

[0032] The present invention is also a partial ribo-substitution method for determining the sequence of a nucleic acid molecule. This method comprises synthesizing a nucleic acid molecule, either *de novo* from a promoter or by extending a primer annealed to the template molecule in four separate reactions. The reactions each have, either four dNTPs and a portion of an rNTP or four 2'-F-NTPs and a portion of an rNTP, or four different non-canonical nucleoside triphosphates, wherein these nucleoside triphosphates have substituents different than a hydroxyl group at the 2' position of the ribose and which the mutant polymerase can use as substrates for synthesis in nucleic acids, and a portion of an rNTP. The reaction products are then cleaved at sites containing an incorporated rNTP by using an alkaline solution or an RNase, and the cleaved nucleic acid fragments are separated according to size so that the sequence of the template molecule may be determined.

[0033] The present invention is also embodiments of a partial ribo-substitution method wherein the nucleic acid synthesis reactions of said method occur in the same reaction mixture and are also part of or linked to a method for nucleic acid amplification, including, but not limited to, NASBA, 3SR, TMA, or other similar methods.

[0034] In still other embodiments of the present invention, the products of either 1, 2, 3, or 4 of the dideoxy-sequencing reactions or of the partial ribo-substitution sequencing reactions are performed or analyzed to determine the presence or absence of a particular nucleic acid, or its relatedness to another nucleic acid, or whether it contains a mutation compared to another nucleic acid.

[0035] The present invention also provides a kit for performing any of the above-identified methods.

[0036] It is an object of the present invention to provide a double-mutant polymerase capable of altered discrimination between canonical and non-canonical nucleoside triphosphates.

[0037] It is an object of the present invention to provide an improved DNA sequencing method.

[0038] It is an object of the present invention to provide a method to detect the presence of a nucleic acid.

[0039] It is an object of the present invention to provide a method to detect the identity of a nucleic acid.

[0040] It is an object of the present invention to provide a method to detect mutations in a nucleic acid.

[0041] It is an object of the present invention to minimize the steps involved in amplifying and sequencing, detecting, identifying and detecting mutations in nucleic acids.

[0042] It is another object of the present invention to provide a method for synthesizing nucleic acid molecules with altered nuclease susceptibility.

[0043] It is another object of the present invention to provide a method for synthesizing nucleic acid molecules comprising at least one non-canonical nucleoside triphosphate.

[0044] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although suitable methods and materials for the practice or testing of the present invention are described below, other methods and materials similar or equivalent to those described herein, which are well known in the art, can also be used.

[0045] Other objects, features and advantages of the present invention will become apparent after examination of the specification, claims and drawings.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0046] Fig. 1 diagrams the transcription products produced by Y639F and W.T. T7 RNAP in the presence of various combinations of rNTPs and dNTPs.

[0047] Fig. 2 shows the effect of dGTP substitution on transcription by the W.T. and Y639F polymerase.

[0048] Fig. 3 shows the effects of single-strandedness and sequence in the initially transcribed region on the activity of Y639F in reactions with 4 rNTPs or 4 dNTPs.

[0049] Fig. 4 shows transcription by Y639F and W.T. polymerase with dGTP or rGTP on poly(dI)·poly (dC).

[0050] Fig. 5 shows primed synthesis of DNA and RNA with Y639F and the W.T. polymerase.

[0051] Fig. 6 shows relative elongation rates of Y639F in "4 rNTP" and "3 rNTP + 1 dNTP" reactions.

[0052] Fig. 7 is a diagram of the mutagenesis strategy involved in creating a mutant SP6 RNA polymerase.

[0053] Fig. 8. is a 3-D visualization of the double-mutant RNAP. The view is of the templating base, 3'-rNMP:template base-pair, and Y639 and H784 side-chains as seen in the structure (Cheetham, et al., Science, 286:2305-2308, 1999) or a T7 RNAP initial transcription complex (PDB 1QLN).

[0054] Fig. 9. discloses incorporation of non-canonical nucleotides by wild-type, Y639F, and Y639F/H784A enzymes. The NTPs (at 0.5 mM) present in each reaction are specified over each gel lane. A: Bgl II-cut pPK5 (Mentesanas, P.E., et al., J. Mol. Biol., 302:1049-1062, 2000) as template. Lanes 1-4: Wild-Type (W.T.); lanes 5-8: Y639F; lanes 9-12: Y639F/H784A; B: Bgl II-cut pPK5 as template. Lanes 1-4: W.T.; lanes 5-8: Y639F; lanes 9-12: Y639F/H784A; C: *Hind*III-cut pT75 (Tabor, S. and Richardson, Proc. Natl. Acad. Sci. USA 82:1074-1078, 1985) as template. Lanes 1-4: W.T.; lanes 5-8: Y639F; lanes 9-12: Y639F/H784A. D: *Hind*III-cut pT75 as template. Lanes 1-4: W.T.; lanes 5-8: Y639F; lanes 9-12: Y639F/H784A.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0055] By "double-mutant polymerase" is meant aT7-type nucleic acid polymerase which has an altered amino acid at the Y639 and H784 positions relative to T7 polymerase compared to the corresponding wild-type polymerase, wherein said mutation or alteration results in the mutant polymerase having a reduced discrimination between non-canonical and canonical nucleoside triphosphates as substrates.

[0056] The term "template" refers to a macromolecular pattern or mold for the synthesis of another macromolecule, composed of a sequence of nucleotides, either rNTPs or dNTPs, that serves to specify the nucleotide sequence of another structure.

[0057] "Nucleotide" refers to a base-sugar-phosphate compound. Nucleotides are the monomeric subunits of both types of nucleic acid polymers, RNA and DNA. "Nucleotide" refers to ribonucleoside triphosphates, rATP, rGTP, rUTP and rCTP, and deoxyribonucleoside triphosphates, such as dATP, dGTP, dTTP, dCTP.

[0058] As used herein, "nucleoside" refers to a base-sugar combination without a phosphate group. "Base" refers to the nitrogen-containing base, for example adenine (A), cytidine (C), guanine (G) and thymine (T) and uracil (U).

[0059] "Incorporation" refers to becoming a part of a nucleic acid polymer. There is a known flexibility in the terminology about incorporation of nucleic acid precursors. For example, the nucleotide dGTP is a deoxyribonucleoside triphosphate. Upon incorporation into DNA, it becomes a dGMP, or deoxyguanosine monophosphate moiety. Although there is no dGTP molecule in DNA, one may say that one incorporates dGTP into DNA.

[0060] As defined herein, a "canonical" nucleoside triphosphate for an RNA polymerase ("RNAP") consists of any ribonucleoside-5'-triphosphate ("rNTP" or "NTP") which has an hydroxyl group at the 2'-position of the sugar, including, but not limited to, the four common ribose-containing substrates for an RNA polymerase -ATP, CTP, GTP and UTP. A 2'-deoxyribonucleoside-5'-triphosphate ("dNTP") which has hydrogen at the 2'-position of the sugar, including, but not limited to, the four common deoxyribose-containing substrates (dATP, dCTP, dGTP and dTTP, also known as "TTP") for a DNA polymerase ("DNAP") is defined herein as a "non-canonical" nucleoside-5'-triphosphate or a "non-canonical NTP" or a "non-canonical nucleotide" or a "non-canonical deoxynucleotide" or a "non-canonical triphosphate" or a "non-canonical substrate" for an RNA polymerase. On the other hand, a "canonical" nucleoside triphosphate for a DNAP consists of any dNTP which has a hydrogen at the 2'-position of the sugar, while an rNTP is defined as a "non-canonical NTP" or a "non-canonical nucleotide" or a "non-canonical substrate" for a DNAP. The terms "canonical" and "non-canonical" are meant to be used herein only with reference to the 2' position of the sugar. Thus, as defined herein, 2',3'-dideoxynucleoside-5'-triphosphates ("2',3'-ddNTPs" or "ddNTPs") are "non-canonical" substrates for an RNAP, but are defined as "canonical" for a DNAP. Further, any other substituent than an hydroxyl group at the 2'-position of ribose or a hydrogen at the 2'-position of deoxyribose, including, but not limited to, a fluorine ("F" or "fluoro" group) or an amino group, would be defined as "non-canonical" for both RNAPs and DNAPs herein. The terms "canonical" or "non-canonical" also are not meant to refer to the nucleic acid bases attached to the sugar moieties. Thus, for example, other natural or modified nucleic acid bases attached to the 1'-position of ribose-5'-triphosphate would still be defined as "canonical" herein.

[0061] By "a (double-mutant) nucleic acid polymerase (enzyme) with reduced discrimination between canonical and non-canonical nucleoside triphosphate substrates", a specific quantitative definition has been calculated as follows to better define the phrase:

[0062] 1. One first determines the K_m and the k_{cat} for each enzyme (mutant and wild-type) using the non-canonical nucleotide as a substrate and using the canonical nucleotide as a substrate, as was described previously (Patra, *et al.*, 1992). The value " K_m " expresses how readily the enzyme will bind the substrate (a larger K_m implies weaker binding) and " k_{cat} " expresses the rapidity with which the substrate, once bound by the enzyme, is reacted upon.

[0063] 2. One next calculates the numerical value for k_{cat}/K_m for each enzyme and for each substrate. By broad scientific consensus, the specificity of an enzyme for a substrate is felt to be most suitably expressed by this ratio.

[0064] 3. For each enzyme, one then calculates the numerical value obtained by using the value of k_{cat}/K_m for the canonical substrate in the numerator and k_{cat}/K_m for the non-canonical substrate in the denominator. This number indicates how much a given enzyme discriminates between the two substrates. For example, if this value equals 1, then the enzyme uses both the canonical and the non-canonical substrates equally well; it does not discriminate between the two substrates. If this value is greater than 1, then the enzyme discriminates by that factor between the two substrates; for example, if the value is 100, then the enzyme discriminates by a factor of 100 in favor of the canonical substrate compared to the non-canonical substrate. Similarly, a value less than 1 means that the enzyme discriminates in favor of the non-canonical substrate over the canonical substrate.

[0065] 4. Finally, one compares the numerical value calculated in step 3 above for the wild-type (W.T.) enzyme with the value calculated in step 3 for the mutant enzyme. If the value calculated for the mutant enzyme is less than the value calculated for the wild-type enzyme, then the mutant enzyme "has a reduced discrimination for the non-canonical substrate compared to the wild-type enzyme." For example, if the values calculated in step 3 are 100 for the wild-type enzyme and 10 for the mutant enzyme, then the discrimination of the mutant enzyme in favor of the canonical substrate (or against the non-canonical substrate) is reduced 10-fold compared to the discrimination of the wild-type enzyme.

[0066] In accordance with the present invention, it has been found that for wild-type T7 RNAP, the average of the k_{cat}/K_m values for the four common rNTPs (ATP, CTP, GTP & UTP) is about 120-fold larger than the average k_{cat}/K_m values for the four common dNTPs (dATP, dCTP, dGTP and dTTP); i.e., the wild-type enzyme discriminates by a factor of 120 for rNTPs vs. dNTPs. For the Y639F mutant enzyme, the average of the k_{cat}/K_m values for the four common rNTPs is only about 6-fold larger than the average k_{cat}/K_m values for the four common dNTPs. Thus, using the average k_{cat}/K_m values for these substrates, the Y639F mutant T7 RNAP enzyme has about a 20-fold reduced discrimination between dNTPs and rNTPs. However, it is

recognized that the difference in discrimination between wild-type and mutant enzymes will vary depending on the non-canonical substrates and the mutant enzymes used. Therefore, for the purposes of this invention, defined herein is "a (mutant) nucleic acid polymerase (enzyme) with reduced discrimination between canonical and non-canonical nucleoside triphosphate substrates" as "a polymerase which has at least a 10-fold reduced discrimination compared to the corresponding wild-type enzyme for non-canonical nucleotides compared to canonical nucleotides, wherein the respective values for discrimination between canonical and non-canonical substrates is calculated using the average of the k_{cat}/K_m values for all four rNTPs and all four dNTPs."

[0067] As used herein "T7-type RNA polymerases" means T7, T3, ϕ I, ϕ IIH, W31, ghl, Y, A1122, SP6 and mitochondrial RNAPs.

In General

[0068] There are many reasons to synthesize nucleic acid molecules containing at least one non-canonical nucleotide. For example, incorporation of a non-canonical nucleotide may make the synthetic nucleic acid more resistant and therefore, more stable, to a nuclease, such as a ribonuclease. Also, one may wish to incorporate one or more non-canonical nucleotides which, for example, change the nuclease digestion pattern so that the product nucleic acid is easier to detect or characterize. For example, because RNase A cleaves RNA only after C or U, replacement of one or both of these rNMPs by a dNMP or other non-canonical nucleotide that is resistant to cleavage by RNase A would alter the digestion pattern of the nucleic acid.

[0069] There are many uses for nucleic acids which have one or more of these properties, such as nuclease resistance. For example, nucleic acids containing at least one non-canonical nucleotide may have advantages for use as ribozymes, or as nucleic acid molecules used for gene therapy, in a vaccine, in an antiviral composition, in an antimicrobial composition, in an anti-sense composition for regulating gene expression, in a composition for hybridization to a complementary nucleic acid, including as a primer, or as a probe for detection of a complementary nucleic acid for a variety of purposes.

[0070] Some nucleic acid molecules, such as those of mixed dNMP/rNMP composition, are highly useful for certain applications, but are presently difficult or impossible to produce on a practical scale. Thus, improved methods for synthesizing such nucleic acid molecules *in vitro* would be highly desirable. For example, probes of mixed DNA-RNA-DNA composition for the Cycling Probe Assay (Duck, P.G., *et al.*, 1990) are currently made using difficult chemical methods.

[0071] Described herein are previously-unknown properties of T7-type RNA polymerases having a non-wild-type amino acid at specific positions within the polypeptides. It has been found that altering the amino acid at these specific positions results in mutant polymerases having at least a 10-fold reduced discrimination between 2'-deoxyribonucleoside-5'-triphosphates (dNTPs) and ribonucleoside-5'-triphosphates (rNTPs) as substrates in *in vitro* nucleic acid synthesis reactions compared to the corresponding wild-type enzymes. Also, it was found that these mutant polymerases also have reduced discrimination for other non-canonical nucleoside triphosphate (NTP) substrates, including 2',3'-dideoxy-ribonucleoside-5'-triphosphates (ddNTPs) and 2'-fluoro-nucleoside-5'-triphosphates (2'-F-NTPs, 2'-Azido-NTPs, 2'-OMe-NTPs). Based on knowledge of these novel properties, methods for using these mutant polymerases for producing nucleic acid molecules containing at least one non-canonical nucleotide and for characterizing nucleic acid molecules by synthesizing nucleic acid molecules containing at least one non-canonical nucleotide *in vitro* are disclosed.

[0072] In one preferred embodiment, the invention uses double-mutant RNA polymerases that efficiently utilize a variety of non-canonical nucleoside triphosphates as substrates.

Mutant Polymerases of the Present Invention

[0073] RNAPs that incorporate non-canonical nucleotides are useful for making RNA molecules for a variety of applications, including, but not limited to, applications previously described herein. In one preferred embodiment, mutant RNAPs are used for synthesis of RNA molecules containing non-canonical nucleotides for use as probes for nucleic acid arrays or for using in silencing genes *in vivo* by means of acting as an interference RNA or RNAi, including, but not limited to, a small interference RNA or siRNA. In some embodiments, the synthesized nucleic acids with non-canonical nucleotides are resistant to non-specific nucleases or have other special properties.

[0074] In another preferred embodiment, double-mutant T7-like RNAPs can be used to synthesize RNA containing non-canonical nucleotides for use in a method referred to "immuno-detection amplified by T7 RNA polymerase" or "IDAT" (Zhang, H.-T., et al., "Protein quantification from complex protein mixtures using a proteomics methodology with single-cell resolution," Proc. Natl. Acad. Sci. USA 98:5497-5502, 2001). In still another preferred embodiment, mutant T7-like RNAPs are used to synthesize RNA containing non-canonical nucleotides in order to analyze specific binding of the nucleic acid to cellular proteins, such as receptors, drug target candidates, drug candidates, or other biomolecules for use in proteomic, pharmacogenomic, or drug screening research. In certain aspects of proteomics research it is

particularly desirable to use a mutant RNAP that can incorporate non-canonical nucleotides which are resistant to certain nucleases and/or which have properties such as having a chemical moiety which can be more easily linked to proteins or other bio-molecules, drug target candidates, or drug candidates which are in close proximity in the cell or when non-covalently bound to the nucleic acid, directly or indirectly, on a nucleic acid or protein array. By way of example, but not of limitation, an azido group in the 2'-position of a non-canonical nucleotide provides methods for linkage of a nucleic acid which contains it to other nucleic acids, proteins, bio-molecules, drug target candidates, or drug candidates in close proximity. For example, such a group can be used to link the nucleic acid to other bio-molecules by chemical or photochemical means.

[0075] As disclosed herein, when the tyrosine in the conserved motif with the consensus sequence ...K-----YG... of T7-like RNAPs is mutated to phenylalanine, the mutant RNAP, including, but not limited to the Y639F mutant of T7 RNAP, the Y631F mutant of SP6 RNAP, or the Y573F mutant of T3 RNAP, readily utilizes 2'-deoxy, 2'-NH₂, and 2'-F NTPs as substrates. However, these mutant RNAPs do not readily utilize non-canonical nucleoside triphosphates with bulkier 2'-substituents or facilitate incorporation of nucleoside triphosphates with modifications at other positions.

[0076] Disclosed herein is an introduction of a second mutation in a T7-like RNAP which alters the conserved histidine corresponding to H784 of T7 RNA polymerase, H779 of SP6 RNAP, or H785 of T3 RNAP, or the terminal histidine in a conserved amino acid sequence, which conserved sequence comprises at least 2, 3, 4, 5 or 6 of the amino acids in the sequence "APNFVH," of these or other T7-like RNAPs, markedly enhances utilization of non-canonical NTPs with bulky 2'-substituents, such as, but not limited to, 2'-OMe and 2'-N₃ (i.e. "azido") groups.

[0077] In preferred embodiments of the invention, the corresponding histidine is mutated to alanine, but the invention also includes embodiments in which the histidine is mutated to another amino acid that occupies less space than the side-chain of histidine and which permits synthesis of RNA which contains non-canonical nucleotides with bulky 2'-substituents, such as, but not limited to, 2'-OMe and 2'-azido groups. Other preferred embodiments of the invention use mutant RNAP enzymes with a second mutation which converts the histidine corresponding to H784 of T7 RNAP or the like histidine of another T7-like RNAP, as described above, so as to also permit improved synthesis of nucleic acids using non-canonical nucleotides, such as, but not limited to, dNTPs or other non-canonical nucleotides which also have modifications at other positions on the nucleic acid base or the sugar moiety and/or NTPs with modifications on the

nucleic acid base or the sugar moiety at positions other than the 2'-position. By way of example, but not of limitation, the modifications on the nucleic acid base or sugar moiety of the nucleotides used in these embodiments can be a fluorescent moiety, such as, but not limited to, a Cy5 dye which is covalently linked to a cytosine or uridine base via an aminoallyl linkage, including, for example, FluoroLink™ Cy5-dCTP or FluoroLink™ Cy5-UTP (Amersham Biosciences).

[0078] Without limiting the invention, it is believed that the Y639F mutation of T7 RNAP does not overcome the barrier to synthesis of nucleic acids with bulkier substituents at the 2'-position of non-canonical nucleotides because the tyrosine of Y639 is positioned so as to discriminate the structure of the substrate NTP, but is not involved in sensing the structure of the 3'-rNMP of the RNA. The crystal structure of a T7 RNAP transcription complex (Fig. 8) reveals that the side-chain of histidine 784 occupies the minor groove side of the 3'-rNMP:template base-pair (Cheetham, G.M.T., and Steitz, T.A., "Structure of a transcribing T7 RNA polymerase initiation complex," *Science* 286:2305-2308, 1999). Though the H784 histidine is well-conserved in T7-like phage RNAPs, it was previously found that an H784A mutant had near wild-type activity, but exhibited enhanced extension of RNAs with mispaired 3'-termini (Huang, J., *et al.*, "Misincorporation by wild-type and mutant T7 RNAPs," *Biochemistry* 39:11571-11580, 2000).

[0079] The present invention discloses methods for using RNAPs having a second mutation, which mutation comprises or corresponds to the H784A mutation of T7 RNAP, to synthesize nucleic acid transcripts containing bulkier non-canonical NMPs. Without limiting the invention, it appears that replacement of histidine, with its bulky side-chain, by alanine makes room for non-canonical nucleotides with bulkier minor groove substituents.

Methods of the Present Invention

[0080] In one embodiment, the present invention is a method for using a double-mutant polymerase for synthesizing *in vitro* a nucleic acid molecule which comprises at least one non-canonical nucleotide in place of at least a portion of the canonical nucleotides. The method comprises the steps of incubating a template nucleic acid in a reaction mixture containing a double-mutant nucleic acid polymerase which has reduced discrimination between canonical and non-canonical nucleoside triphosphates, including between dNTPs and rNTPs, and the appropriate canonical or non-canonical nucleoside triphosphates which are substrates for the nucleic acid polymerase. One then follows standard polymerase reaction protocols and creates the synthesized nucleic acid molecule.

[0081] Preferably, the reactions also contain inorganic pyrophosphatase, which is known to increase the yields in *in vitro* transcription reactions (Cunningham, P.R. and Ofengand, J., 1990) and to reduce pyrophosphorolysis in *in vitro* DNA synthesis reactions (Tabor, S., and Richardson, C.C., 1990), as well as buffers and other components which are known to those of skill in the art to be optimal for the particular W.T. polymerase used. Cunningham and Ofengand (1990) provide an example of conditions which may be used for unprimed synthesis with T7 RNA polymerase or mutant T7 RNAPs, although one of skill in the art will recognize, with respect to reactions with these enzymes or other enzymes, the need to optimize the concentrations and ratios of canonical and non-canonical NTP substrates according to the respective K_m and application and to modify reaction conditions, such as temperature, amount of enzyme, salt concentration, or divalent cation (e.g., Mg^{2+} or Mn^{2+}) concentration, in order to produce improved results such as higher yield or a greater percentage of full-length products.

[0082] In a preferred form of this method, the resulting synthesized nucleic acid molecule has a different susceptibility to a nuclease compared to a nucleic acid synthesized by the corresponding non-mutant nucleic acid polymerase under identical reaction conditions with canonical substrates. By "different susceptibility" one means reduced, increased, or, in the case of synthetic nucleic acids containing both canonical and non-canonical nucleotides, altered susceptibility to a nuclease, which may be either a DNase or RNase. The nature of the reduced, increased or altered susceptibility to a nuclease is also related to the properties of the nuclease. For example, a nucleic acid resistant to RNase A, which cleaves RNA only after C or U, may be synthesized using fewer non-canonical nucleotides (e.g., dNTPs or 2'-F-NTPs) than a nucleic acid which is resistant to RNase I, which cleaves after every base.

[0083] In a preferred form of the present invention, the resulting synthesized nucleic acid is a ribozyme or a nucleic acid molecule used for gene therapy, in a vaccine as an antiviral composition, in an antimicrobial composition, as an antisense composition for regulating gene expression, in a composition for hybridization to a complementary nucleic acid, such as for a primer, or as a probe for detection of a complementary nucleic acid.

[0084] The resulting synthesized nucleic acid may be either single- or double-stranded.

[0085] The present invention is also a kit for performing the method of synthesizing a nucleic acid containing at least one non-canonical nucleotide. Typically, the kit contains the double-mutant nucleic acid polymerase with a reduced discrimination for non-canonical compared to canonical substrates and data or information describing conditions under which the method may be performed.

[0086] The present invention is also improved methods for sequencing nucleic acids using a mutant nucleic acid polymerase of the present invention.

[0087] Because 2',3'-dideoxynucleotides are not substrates for wild-type RNA polymerase, it previously has not been possible to use the Sanger method for determining the sequence of a nucleic acid with an RNA polymerase, although 3'-deoxy- or 3'-hydroxymethyl analogs have been used as terminators for Sanger-like sequencing with RNA polymerases.

[0088] However, 2',3'-ddNTPs are substrates for the mutant nucleic acid polymerases of this invention which can also utilize both rNTPs and dNTPs as substrates, and the present invention is also a method for sequencing nucleic acid molecules (DNA or RNA) using the double mutant nucleic acid polymerase and 2',3'-ddNTPs as terminators.

[0089] In one embodiment of this method, the nucleic acid to be sequenced, whether DNA or RNA, is used as a template for *in vitro* nucleic acid synthesis from a primer (i.e., primed synthesis) using the double-mutant RNA polymerase which has a reduced discrimination for dNTPs compared to rNTPs. Each of four different reactions also contains an amount of at least one nucleoside triphosphate corresponding to each nucleic acid base represented in either DNA or RNA, chosen from among the 2'-deoxynucleotides dATP, dCTP, dGTP and dTTP or dUTP, or the four common ribonucleotides ATP, CTP, GTP and UTP, or the 2'-fluorine-substituted nucleotides 2'-F-ATP, 2'-F-CTP, 2'-F-GTP and 2'-F-UTP or 2'-F-TTP. A 2',3'-dideoxynucleotide is also included in each *in vitro* nucleic acid synthesis reaction in an amount that will result in random substitution by the dideoxynucleotide of a small percentage of the corresponding rNTP, dNTP or 2'-F-NTP that is present in the reaction and that would be incorporated into the synthetic nucleic acid in a template-dependent fashion.

[0090] Thus, each DNA synthesis reaction yields a mixture of DNA fragments of different lengths corresponding to chain termination wherever the dideoxynucleotide was incorporated in place of the ribo-, 2'-deoxy- or 2'-fluoro- nucleotide. The DNA fragments are labeled, either radioactively or non-radioactively, by one of several methods known in the art and the label(s) may be incorporated into the DNA by extension of a labeled primer, or by incorporation of a labeled ribo-, deoxy-, 2'-fluoro- or 2',3'-dideoxy- nucleotide.

[0091] By carrying out DNA synthesis reactions for each of the dideoxynucleotides (ddATP, ddCTP, ddGTP and ddTTP or ddUTP), then separating the products of each reaction in adjacent lanes of a denaturing polyacrylamide gel or in the same lane of the gel if different distinguishable labels are used for each separate reaction, and then detecting those products by one of several radioactive or non-radioactive methods known in the art, the sequence of the DNA template can be read directly. Also, other matrices than polyacrylamide which separate the

fragments based on size may be used. Those with skill in the art will also recognize that other nucleotide analogs generally used for reducing sequencing compressions, such as ribo-, deoxy- or 2'-fluoro- nucleoside triphosphates containing 7-deaza-guanine or inosine, may also be used in place of the ribo-, deoxy- or 2'-fluoro- nucleotide for which the respective analog is used.

[0092] The present invention also comprises another embodiment of this method for sequencing using 2',3'-ddNTPs and a mutant RNA polymerase which has a reduced discrimination for dNTPs compared to rNTPs. In this embodiment of the method, the nucleic acid to be sequenced, whether DNA or RNA, is used as a template for *de novo* (i.e., unprimed) *in vitro* nucleic acid synthesis beginning at an RNA polymerase promoter. In this embodiment, a primer is omitted from the reactions and depending on the promoter sequence, in addition to the other components used in the first embodiment, an amount of a dinucleoside tetraphosphate or a trinucleoside penta-phosphate may be added as an initiating nucleotide (Moroney and Piccirilli, 1991) so that the majority of nucleic acid synthesis is initiated from a single site. Because no primer is used, the labeling of the sequencing products must be carried out by one of the other methods envisioned and discussed with respect to the first embodiment of the method or by incorporating a label into or on an initiating dinucleotide or trinucleotide. In other respects, the second embodiment of this sequencing method of the invention is similar to the first embodiment of the method.

[0093] The present invention also comprises methods of sequencing nucleic acids by partial ribo-substitution using mutant nucleic acid polymerase which have a reduced discrimination between non-canonical and canonical nucleotides. These methods have advantages over the partial ribo-substitution sequencing method described by Barnes (1977), which relied on the use of a Mn^{2+} -containing reaction buffer to relax the ability of a wild-type DNA polymerase to discriminate between dNTPs and rNTPs. Further, only DNA was used as a template for this method and *de novo* (i.e., unprimed) nucleic acid synthesis was not envisioned. In contrast, the ribo-substitution sequencing method of the present invention uses a mutant nucleic acid polymerase which has an inherent reduced discrimination between dNTPs and rNTPs and, although it may be included, Mn^{2+} is not required in the sequencing reactions. In still another embodiment, 2'-fluorine-substituted NTPs are used in place of dNTPs in the ribo-substitution reaction. Embodiments of the method of the present invention also include sequencing using either DNA or RNA as templates and sequencing using either nucleic acid primers or *de novo* nucleic acid synthesis from an RNA polymerase promoter sequence.

[0094] Since incorporation of the sequence-delimiting ribonucleotide does not terminate nucleic acid synthesis during partial ribo-substitution sequencing, all of the radioactive or non-

radioactive label must be incorporated into the sequencing reaction products prior to incorporation of the first ribonucleotide in order to avoid multiple labeled products produced from each nucleic acid molecule synthesized. Multiple labeled products, starting from different positions on the template, would make it difficult or impossible to interpret sequence results. Thus, labeling of nucleic acid products for partial ribo-substitution sequencing must be accomplished by means such as incorporating the label into a primer, when used, or into an initiating di- or tri-nucleotide, or by prelabeling in the presence of an amount of a labeled nucleotide which will be used up or destroyed and/or limiting 2'-deoxy- or 2'-fluoro- nucleotides prior to addition of the sequence-delimiting rNTPs.

[0095] The present application envisions sequencing reactions wherein all of the distinguishable non-radioactive labels in, or attached to, or connected with, the products from more than one of the reactions, up to and including all four of the sequencing reactions for a template, or even reactions from multiple templates if distinguishable non-radioactive labels are present, are detected in or from a single lane of a sequencing gel or a single capillary electrophoresis tube or a single matrix or means of any kind using an automated sequencer or other detection device.

EXAMPLES

Example 1: Creation and Characterization of a Mutant T7 RNA Polymerase

A. Materials and Methods

[0096] Nucleic acids and NTPs: Nucleotides were from Pharmacia or USB/Amersham. Polynucleotides were from Pharmacia and the Midland Certified Reagent Company. Synthetic DNAs were prepared at the UTHSCSA DNA synthesis facility on an Applied Biosystems DNA synthesizer and purified by HPLC. A synthetic RNA 12mer was from the Midland Certified Reagent Company. Plasmids pT75 (Tabor and Richardson, 1985) and pBS (Stratagene Inc.) were purified from *E. coli* by alkaline-lysis and cesium chloride gradient centrifugation (Sambrook, et al., 1989). Radioactive nucleotides were from NEN Dupont or ICN.

[0097] Preparation and purification of mutant polymerases: Construction, expression, and purification of the T7 RNAP mutants was described previously (Bonner, et al., 1992).

[0098] Transcription reactions: Transcription reactions were carried out in 40 mM Tris-Cl pH 8.0, 15 mM MgCl₂, and 5 mM DTT or 20 mM Manganese Citrate pH 8.0, 5 mM DTT at 37°C. Template, polymerase, and NTP concentrations were as indicated in the legends to the figures and tables. Relative activity determinations were made by taking 4 µl aliquots of

reactions at 5, 10, and 20 minute time points and spotting on to DE81 filter paper.

Unincorporated nucleotides were separated from incorporated nucleotides by washing the filter paper with .5 M KH_2PO_4 pH 7.0 and retained radioactive nucleotide was quantitated with a Molecular Dynamics phosphorimager. Radioactive NTPs used were as indicated in figure and table legends. To evaluate rNTP/dNTP selectivity, reactions were run with 4 rNTPs and pT75 as template and $\alpha\text{-P}^{32}$ rNTPs or $\alpha\text{-P}^{32}$ dNTPs were used to label the transcripts. The relative rate of incorporation of an rNTP vs. its cognate dNTP was determined from the relative percentages of labeled rNTP vs. dNTP incorporated into DE81 retainable RNA at 5, 10, and 20 minute time points. Apparent miscoding frequencies were determined similarly, though in this instance the template was a single-stranded homopolymer and the reaction contained the complementary unlabeled rNTP, and complementary $\alpha\text{-P}^{32}$ rNTP or one of the 3 non-complementary $\alpha\text{-P}^{32}$ rNTPs. The relative percentages of labeled complementary vs. non-complementary rNTPs incorporated at 5, 10 and 20 minute time points gave an apparent miscoding rate. The rNTP/dNTP selectivity assay was used to test the following T7 RNAP mutants for effects on substrate discrimination: D537S, D537E, S539A, R551S, D552S, R627S, K631S, L637A, Y639S, Y639F, Y639A, G640A, F644A, G645A, Q649S, !810S, H811S, H811A, D812A, D812E, D812S, D812N, D879E+ Δ F882+ Δ A883, D879E+A881T+ Δ A883, D879E+F884+A885, D879E+F882Y, D879E+ Δ A883, D879E+F882W, D879E.

[0099] Elongation rate determinations were carried out as described (Golomb and Chamberlin, 1974) with some variations (Bonner, et al., 1994).

[00100] Determination of NTP K_m and k_{cat} was as described previously (Patra, et al., 1992).

B. Results

[00101] Structure of the transcripts synthesized by Y639F and the W.T. enzyme with rNTPs and dNTPS: Fig. 1 diagrams the structure of transcription products produced by Y639F and W.T. T7 RNAP in the presence of various combinations of rNTPs and dNTPs. The template was pT75 (Tabor and Richardson, 1985) cut with *Hind*III so that transcription from its T7 promoter generates a 59-base run-off transcript. Electrophoresis was on a 20% polyacrylamide 6M urea gel. Plasmid and polymerases were at concentrations of 10^{-7} M, and NTP concentrations were .5 mM (all rNTPs and dTTP), 1 mM (dATP, dGTP), or 5 mM (dCTP). $\gamma\text{-P}^{32}$ -GTP was added to radiolabel the transcription products. Wild-type (WT) or Y639F mutant (Mu) polymerases and NTPs used are as indicated. Poly-rG products of various sizes are labeled in lane A ("2G", "3G", etc...) and heterogeneous sequence abortive transcripts of different lengths

are indicated by "4H", "5H", etc... in lane c. Lanes q-t are a 10-fold longer exposure of lanes m-p.

[00102] Fig. 1 shows transcription reactions carried out with the W.T. enzyme or the Y639F mutant polymerase and a T7 ϕ -10 promoter template. Transcription by T7 RNA polymerase, like other RNA polymerases, is characterized by an initial, poorly processive 'abortive' phase of transcription during which the short, nascent transcript frequently dissociates from the ternary complex. When the transcript reaches a length of ~9 bases transcription becomes highly processive and the transcript becomes stably associated with the elongation complex. On this promoter, which initiates with GGGAGACCGGAAU (SEQ ID NO:1), T7 RNAP can also synthesize long poly-G ladders (labeled "2G", "3G", etc. in lanes a and b) when rGTP is the sole NTP present (Martin, *et al.*, 1988). Lanes c and d of the electrophoretic gel display the transcription products typical of runoff reactions counting all 4 rNTPs: there are abortive transcripts ranging up to 8 bases in length and a 59-base runoff product. The 4mer length the sequences of the poly-G transcripts and the abortive transcripts made in the presence of all 4 rNTPs (labeled "4H", "5H", etc.) diverge and no longer co-migrate with the poly-G transcripts of equivalent size.

[00103] When rATP is omitted from the transcription reactions (lanes e and f) normal elongation of the initially synthesized GGG trimer cannot occur. There are no heterogenous sequence abortive transcripts or 59 base runoff products made and instead long poly-G transcripts, as in lane a, are made. Adding dATP to reactions lacking rATP does not change the transcripts produced by the W.T. enzyme (lane g). However, with the mutant enzyme it is observed that addition of dATP (lane h) allows synthesis a long runoff transcript as well as synthesis of heterogenous sequence abortive transcripts that do not co-migrate with the poly-G transcripts. Observation of an abortive transcript in lane h running near the position of the "4H" band in lane d confirms extension of the GGG trimer with an A but note that the major 4mer transcript in lane h migrates close to, but not precisely with, the major 4mer in lane d or in adjacent lane i. This is consistent with the expectation that these 4mers will have identical sequence and length but different structure (i.e., rGrGrGrA in lanes d or i; rGrGrGdA in lane h). It should also be noted that some poly-G transcript synthesis is observed in lane h. For example, in lane h it is observed that both a heterogenous sequence 4mer migrating near the "4H" position and a smaller amount of 4mer band migrating at the "4G" position. When 4 rNTPs are present (lanes c or d) the synthesis of poly-G transcripts is more completely suppressed. This indicates that dATP is utilized by Y639F, but not as efficiently as rATP.

[00104] When rCTP is omitted from the reaction, transcripts terminate predominately at the 6mer length because rCMP is normally first incorporated at position 7 (lanes i, j). Addition of dCTP does not allow extension of the 6mer in reactions with the W.T. enzyme (lane k). However, addition of dCTP to reactions with Y639F allows extension beyond the 6mer length and synthesis of the runoff transcript (lane l). Again, the following should be noted: 1. The transcripts larger than 6 bases do not co-migrate with their counterparts in lanes c or d consistent with the expected structural difference despite length and sequence identity, 2. there is more termination at the 6- and 7mer points in lane l than in lanes c or d, indicating that Y639F uses dCTP well, but not as efficiently as it utilizes rCTP.

[00105] In lanes m and n UTP was omitted from the reactions. Lanes q-t show a 10-fold longer exposure of lanes m-p. Within the set of 4 NTPs, UTP is unique on this template since it first becomes incorporated into the transcript at the 13 base position. This corresponds to a transcript length subsequent to the transition from abortive to processive transcription. As a consequence of this transition, the ternary complex becomes more stable (Martin, et al., 1988). Therefore, when transcript extension is blocked during the processive phase of transcription, the stalled ternary complex does not rapidly dissociate (Shi, et al., 1988). Instead it remains stalled on the template, near the promoter, and blocks reinitiation. For this reason a large decrease in the overall amount of transcription when UTP is omitted from the reaction in lanes m and n was observed.

[00106] A longer exposure of lanes m and n (lanes q, r) does, however, reveal transcription products of the expected structure. When TTP is added to these reactions (lanes o, p, s, t) synthesis of the 59 base runoff transcript is observed with both the W.T. and Y639F mutant.

[00107] The ability of the W.T. enzyme to extend transcripts with TTP when it is unable to extend transcripts with the other dNTPs is also likely to be related to unique position at which UTP/TTP first becomes incorporated into the transcript. The amount of transcript termination or extension that occurs with a particular dNTP depends simply on the relative rates of ternary complex dissociation or dNMP incorporation (McClure and Chow, 1980). During abortive transcription complex, dissociation must be more rapid than the rate at which the W.T. enzyme incorporates dNMPs into its transcripts. During processive transcription even the expected slow rate of dNMP incorporation by the W.T. enzyme must be competitive with the slow rate of dissociation of the stable elongation complex, and an ability of the W.T. enzyme to incorporate dNMPs becomes manifest. Because elongation during the processive phase of transcription is fast (~230 bases/sec, Golomb and Chamberlin, 1974), while initiation and progression through abortive transcription is slow (Martin and Coleman, 1987; Martin, et al., 1988), elongation of the

transcript from 11 to 59 bases is expected to contribute less than 10% to the time required for transcript synthesis. As a consequence the phase of transcription during which TTP is incorporated may not be the rate limiting step in synthesis of the runoff transcript even for the W.T. enzyme. These considerations imply that one need not expect to see marked differences in synthesis of a relatively short runoff transcript by Y639F or the W.T. enzyme when TTP is substituted for UTP.

[00108] Lanes u and v show reactions in which two rNTPs have been substituted with dNTPs (dATP and dCTP), and lanes w and x show reactions in which three rNTPs have been substituted with dNTPs (dATP, dCTP, and TTP). Synthesis of the 59 base runoff transcript is observed with the mutant polymerase (lanes v and x) indicating that Y639F can carry out synthesis even when 3 rNTPs are substituted with dNTPs.

[00109] While the abortive transcript patterns in lanes v and x may largely be described as a combination of the patterns observed in lanes h and l there are some important distinctions. The transcript patterns in lanes v and x reveal that the structure of the transcript, as well as the structure of the NTP, affect the rate at which NMPs are added to the transcript. For example, in lanes v and x there is an increase in termination at the 6mer length relative to lane l (the 6mer in lanes v or x runs slightly slower than the poly-G 5mer in the adjacent lane). In lanes l, v and x synthesis of the 7mer involves incorporation of dCMP but in lanes v and x the 6mer contains a dAMP at its 3' end and in lane l it contains an rAMP.

[00110] The increase in termination at the 6mer point in lanes v or x indicates that the presence of a dNMP on the 3' end of the transcript reduces the ability of the polymerase to further incorporate NMPs. The high level of termination after synthesis of the rGrGrGdA 4mer in lane h relative to the observed termination after synthesis of the rGrGrGrA 4mer in lanes c or d similarly indicates that the presence of deoxynucleotides in the transcript, at least at the 3'-end, influences subsequent extension of the transcript.

[00111] Fig. 2 shows the effect of dGTP substitution on transcription by the W.T. and Y639F polymerase. Polymerases, template (supercoiled(Sc) or *Hind*III linearized (Li) pT75), and NTPs were as indicated. Polymerase and template concentrations and electrophoresis conditions as in Fig. 1. Left panel: Labeling was with α -P³² rGTP (a, d, g, j) or α -P³² dGTP(all other lanes). The runoff transcript from the *Hind*III-cut template is indicated in lane f. Right panel: Labeling was with α -P³²-rGTP (a, b, f, g) or α -P³²-dGTP (all other lanes). Poly-rG and poly-dG products of various sizes are indicated in lanes a, c, d. Alignment of these transcript patterns with those in lanes b, c, h, and i in the left panel reveals that the added complexity of the transcript pattern in the latter set of lanes is due to the presence of a mixture of heterogeneous

sequence and poly-G transcripts. Heterogenous sequence abortive transcripts are indicated in lane c of the left panel ("4H", "5H").

[00112] Fig. 2 reveals the effects of substituting dGTP for rGTP in transcription reactions with the W.T. or Y639F polymerase. In reactions containing only dGTP and *Hind*III- cut pT75 as the template, the mutant polymerase synthesizes poly-dG transcripts up to 4 bases in length (lane c, right panel). Note that--consistent with the assumed structural differences--the poly-dG transcripts ("2dG", "3dG", etc.) in lane c do not co-migrate with the poly-rG transcripts ("2rG", "3rG", etc.) in lane a despite length and sequence identity. When a supercoiled template is used, poly-dG transcripts up to 5 bases in length are obtained (lane d, right panel). In lane e rGMP is added to reactions which contain only dGTP. Ribo-GMP can serve as the initiating, but not elongating nucleotide during transcription (Martin and Coleman, 1989). With rGMP it was therefore asked whether either polymerase can elongate with dGTP if an rNMP is provided for initiation. Addition of rGMP to reactions with dGTP further extends the lengths of the transcripts obtained with the mutant polymerase (lane e, right panel). With the W.T. enzyme very little synthesis is observed in reactions with dGTP (lanes h-j, right panel), though the normal pattern of poly-rG synthesis is observed in the rGTP reactions (lanes f, g, right panel).

[00113] When reactions contained three rNTPs and dGTP synthesis of runoff transcript from the *Hind*III-cut template is reduced much more than in reactions in which rGTP is present but other ribonucleotides are substituted with deoxynucleotides. For example, there is very little runoff transcript in lane h of the left panel of Fig. 2. Addition of rGMP increases the amount of runoff transcript made by the mutant enzyme in a reaction containing dGTP (lane i) but the amount of runoff transcript is still much less than in reactions with rGTP. On a supercoiled template (lanes a-c, left panel) high levels of long transcripts are obtained with the mutant enzyme in the reactions with dGTP. The W.T. enzyme shows no transcript synthesis in any of the reactions with dGTP irrespective of whether supercoiled templates or rGMP is used.

[00114] Examination of Fig. 2 shows that the marked reduction in runoff transcript synthesis by the mutant enzyme in reactions with dGTP is not due to a deficit in initiation. In fact, in all of the reactions with dGTP abundant synthesis of 2~6 base transcripts with the mutant enzyme was observed. The low level of runoff transcript synthesis means that these short transcripts are being inefficiently extended to greater lengths. It should also be noted that the abortive transcript pattern seen in lanes b, c, h, and i of the left h panel in Fig. 2 is too complex to be accounted for by presuming one product of each base length. Aligning the transcripts produced in the presence of dGTP only with those produced with dGTP, rUTP, rCTP, and rATP allows us to identify the predominant abortive transcripts in lanes c, d, h, and i of the left panel as

poly-dG products. In lane c of the left panel the major non poly-dG abortives are indicated by "4H" and "5H". In lanes b, c, h, and i of the left panel of Fig. 2 a pattern similar to that of lane h in Fig. 1 is observed. The presence of a mixture of normally extended heterogenous sequence and poly-G abortive transcripts indicates that 'normal' transcript extension is inefficient.

[00115] It has been shown that mutations in T7 RNAP that reduce phosphodiester bond formation rates cause poly-rG transcript synthesis even when 4 rNTPs are present (Bonner, et al., 1994). It can be seen that the ratio of poly-G to heterogenous sequence transcripts in lanes b, c, h, i of the left panel of Fig. 2 is greater than in lane h of Fig. 1, indicating a greater deficiency in normal transcript extension when dGTP is substituted for rGTP than when dATP is substituted for dGTP. Note that this occurs even though normal extension of the dGdGdG trimer in lane i of the left panel of Fig. 2 (for example) would involve addition of a ribo-AMP while extension of the rGrGrG trimer in lane h of Fig. 1 would involve addition of a deoxy-AMP, clearly highlighting the role of transcript structure, as well as substrate structure, in determining the efficiency of transcript extension. These results show that Y639F can initiate and elongate transcripts with dGTP substituted for rGTP but normal extension of the transcripts in the 2-8 base range is impaired leading to a large increase in the proportion of poly-dG and short transcripts synthesized. Addition of rGMP to serve as the initiating nucleotide and the use of a supercoiled template both enhance the ability of the mutant to extend the short transcripts during the initial stages of transcription.

[00116] Barriers to initiation and extension of the initial transcript with dNTPs: Fig. 2 reveals that Y639F can efficiently transcribe with dGTP the initial G segment of a promoter that initiates "GGGA" but is severely blocked in extending the dG trimer with the A. This implies that the sequence of the initially transcribed region may influence the efficiency with which Y639F can extend the transcript when using dNTPs. Fig. 2 also shows that supercoiling, which presumably facilitates unwinding of the template, enhances the activity of Y639F when using dNTPs. To evaluate the effects of sequence and single-strandedness in the initially transcribed region on the activity of Y639F when using dNTPs, transcription was examined from a set of synthetic promoters which differed in the sequence of their initially transcribed regions and in being fully double-stranded or single-stranded in their initially transcribed regions (Fig. 3).

[00117] Fig. 3 shows the effects of single-strandedness and sequence in the initially transcribed region on the activity of Y639F in reactions with 4 rNTPs or 4 dNTPs. Poly-rG transcripts of various sizes are indicated in lane a. Reactions contained the indicated NTPs and polymerases. Polymerase and promoter concentrations were 10^{-6} M and 10^{-5} M, respectively. NTP concentrations and electrophoresis as in Fig. 1. Indicated in some of the lanes are the

sequences of different transcripts as deduced from alignment with poly-rG or poly-dG ladders synthesized in the presence of rGTP or dGTP only. The synthetic promoter templates used are double-stranded and have the sequence--CGAAATTAATACGACTCACTATA (SEQ ID NO:2)-in their -23 to -1 regions. The promoters differ in their initially transcribed regions as follows: b-e, t, u: GGACT; f-j, v, w: GAGACCGG; a, j-m, x, y: GGGAGACC; n, o, z: GGAAAATT; p-s: GGGGGGGGGGGGACT (SEQ ID NO:3). The promoters also differ in being double-stranded (b, d, f, h, j, l, p, r) or single-stranded (other lanes) in their transcribed regions.

[00118] For most of the promoters tested, transcription with rNTPs was not markedly affected by having the initially transcribed region be single-stranded. However, when transcribing with 4 dNTPs, Y639F was more active on the partially single-stranded promoters. For example, in lane e (partially single-stranded promoter) of Fig. 3 transcription products are both more abundant and extend to greater lengths than in lane d, where the promoter is fully double-stranded. A similar comparison may be made between lanes m and l or s and r. Regarding the sequence, it was found that a promoter which initiated with 3 G's was superior to a promoter which initiated with two, which was in turn superior to a promoter which initiated with just one G. Thus Y639F activity when using dNTPs was greatest on a promoter which initiates "GGGAGACC" (lanes l and m). The initially transcribed region of this promoter corresponds to the consensus sequence for T7 promoters in the +1 to +6 segment. This was the only promoter which, when fully double stranded, gave rise to high levels of transcript synthesis with Y639F in reactions containing only dNTPs (lane l).

[00119] Promoters which initiated with 2 G's (lanes d, e, o) gave lower levels of transcript synthesis in the 4 dNTP reactions, and a promoter which initiated with 1 G (lanes h, i) was not utilized by Y639F in reactions with 4 dNTPs. Within the initially transcribed region, elements other than the number of G's appear to be important. For example, it was found that the W.T. or Y639F polymerases are less efficient in initial transcription extension on the T7 promoter found in the pBS plasmid which initiates GGGC, than on the ϕ 10 promoter found in pT75 which initiates with the consensus GGGAGA (data not shown).

[00120] Another example is evident in lanes n and o which show the transcripts obtained on a partially single-stranded promoter that initiates GGAAAUU. Like the promoter used in lanes c and e, this promoter initiates with 2 G's but normal transcript extension on this promoter is less efficient than on the promoter that initiates GGACU. In the 4 rNTP reactions (lanes c, n), the proportion of short transcripts (dimers, trimers) is greater in lane n and significant amounts of poly-rG transcripts beyond the dimer length in lane n but not in lane c was observed. In the 4 dNTP reactions almost all of the transcripts in lane o terminate at the trimer or tetramer length.

Because increasing the number of Gs from 1 to 3 enhanced Y639F activity when using dNTPs, a promoter was tested which initiated with a run of 11 G's (lanes p-s). A potential drawback to such a promoter is that such a long run of G's could inhibit the ability of the polymerase to unwind the template. Since T7 promoters with more than 3 consecutive G's in the initial region do not occur naturally, it may be that for other reasons such sequences do not favor initial transcript extension. In fact, it was found that this promoter is a poor template. When it is fully double-stranded, initiation and extension of the transcript is inefficient with either rNTPs (lane p) or dNTPs (lane r), consistent with the expectation that a promoter with this sequence would be difficult to melt. Initiation and transcript extension is enhanced when this promoter is partially single-stranded (lanes q and s), but while poly-G transcripts from 2 to 7 or more bases in length are abundant, runoff transcripts of the expected length are not predominant products. In reactions with 4 rNTPs the transcript patterns of the W.T. and Y639F polymerases are virtually identical to avoid repeating the 4 rNTP reactions with the W.T. enzyme in Fig. 3. Lanes t-z show that the W.T. enzyme is virtually inactive in reactions with 4 dNTPs and the same set of promoter used in lanes b-s.

[00121] Relative selectivity of the mutant and W.T. polymerases for dNTPs and rNTPs: Figs. 1-3 present a qualitative analysis of the structure of the transcripts produced by the W.T. and Y639F polymerases with various combinations of NTPs. They show that Y639F can use dNTPs with high efficiency and that both transcript and substrate structure play a role in determining the efficiency of transcript extension. To obtain a quantitative measure of the relative selectivity of the W.T. and mutant polymerases for dNTPs vs. rNTPs under conditions where transcript structure was not a complicating factor, reactions were carried out in which all 4 rNTPs were present but rNTPs or dNTPs were used to radiolabel the transcripts (see Table I).

Table I: Relative Selectivity of Y639F and W.T. Polymerase for rNTPs vs. dNTPs

	rATP/dATP	rUTP/dTTP	rCTP/dCTP	rGTP/dGTP
Y639F	8.5-10 (Mg) [*] .9-1.0 (Mn) [*]	1.7-1.9 .48-.76	2.4-2.5 .55-.80	.93-1.6 .98-.99
W.T.	72-83 (Mg) [*] 6-14 (Mn) [*]	22-25 2.3-2.8	110-150 4.3-5.1	51-67 4.4-6.3

Reactions were carried out with all 4 rNTPs (.5 mM) in great excess over radioactive rNTPs or dNTPs. Relative selectivity was determined from the relative percentages of radioactive rNTP vs. dNTP incorporated into RNA. Maximal incorporation was less than ~30% of total input radioactivity with all data points used so as to limit effects due to changing NTP concentrations during the experimental time course. The numbers shown give the range for 2 experiments.

^{*}For each rNTP/dNTP and polymerase the upper numbers are those obtained in Mg²⁺ buffer, the lower numbers are from Mn²⁺ buffer. Polymerases at 10⁻⁸ M. Template was supercoiled pT75 at 10⁻⁷ M.

[00122] Under these conditions the unlabeled rNTPs were present in vast excess relative to the labeling NTPs so labeling dNTPs are almost always incorporated adjacent to rNTPs and into transcripts of nearly uniform rNMP structure. On average the mutant enzyme is ~20-fold less selective for rNTPs over dNTPs than the W.T. enzyme. This assay was used to screen our collection of T7 RNAP active site mutants for increased dNTP utilization. In this screen dATP incorporation in transcription reactions was examined with all of these mutants using supercoiled pT75 as the template, 4 cold rNTPs, and P32-rATP or P32-dATP to label. Since these results were negative with the exception of the tyrosine 639 mutants (not presented here), but the mutants tested in this way are listed in "Materials and Methods". It has been reported that use of Mn²⁺ instead of Mg²⁺ decreases substrate discrimination and increases miscoding for a number of polymerases (Tabor and Richardson, 1989; Nivogi and Feldman, 1981). Therefore, the rNTP/dNTP selectivity of the mutant and W.T. enzymes were examined in Mn²⁺-citrate buffer (Tabor and Richardson, 1989). With Mn²⁺ the preference of both the W.T. and Y639F polymerases for rNTPs over dNTPs was markedly reduced.

[00123] Relative activity of the W.T. and Y639F polymerases with different NTP combinations: The relative activities of the Y639F and W.T. polymerases with supercoiled pT75

as a template and various combinations of rNTPs/dNTPs were measured in both Mg^{2+} and Mn^{2+} buffers (see, Table II,).

Table II: Relative Activity of Y639 F and W.T. Polymerase with Different rNTP/dNTP Mixes

	W.T. (Mg^{2+})	Y639F (Mg^{2+})	W.T. (Mn^{2+})	Y639F (Mn^{2+})
4 rNTPs	200	200	21-23	9
3 rNTPs, dTTP	11-13	90-96	7-8	4
3 rNTPs, dUTP	9-11	82-91	10-11	7-8
3 rNTPs, dATP	1-2	73	1-2	2-3
3 rNTPs, dCTP	4-9	86	15	7-11
3 rNTPs, dGTP, rGMP	<.5	95-109	1-3	1.4-2.5
3 rNTPs, dGTP	<.5	43-63	.5-2	3
2 rNTPs, dCTP, dUTP	2-6	27-29	3	4-7
2 rNTPs, dCTP, dTTP	2	30	5-7	5
2 rNTPs, dCTP, dATP	<.5	20-29	.5-.9	4-7
2 rNTPs, dTTP, dGTP, rGMP	<.5	13-15	<.5	3-6
2 rNTPs, dATP, dTTP	<.5	11-14	<.5	3-4
2 rNTPs, dCTP, dGTP, rGMP	<.5	11-14	<.5	2-5
2 rNTPs, dATP, dGTP, rGMP	<.5	5-6	<.5	1-1.5
1 rNTP, dCTP, dATP, dTTP	<.5	12-14	<.5	.7-2
1 rNTP, dCTP, dATP, dUTP	<.5	10-11	<.5	2-3
1 rNTP, dTTP, dCTP, dGTP*	<.5	10-13	<.5	2-4
1 rNTP, dTTP, dATP, dGTP*	<.5	11	<.5	1.5-2
1 rNTP, dCTP, dATP, dGTP*	<.5	3-5	<.5	1-2
4 dNTPs, rGMP	<.5	<.5	<.5	.5

*These reactions also contain rGMP. Numbers give ranges from 2 experiments.

Template was supercoiled pT75 (10^{-7} M), polymerases at 10^{-8} M (in Mg^{2+} buffer) or 10^{-7} M (in Mn^{2+} buffer). rNTPs, rGMP, dTTP were at .5 mM; dATP, dGTP were at 1 mM; dUTP was at 2.5 mM, dCTP was at 5 mM. From top to bottom the labeling NTPs were: α -P³² rGTP, α -P³² rCTP, α -P³² rCTP, α -P³² dATP, α -P³² dCTP, α -P³² dGTP, α -P³² dGTP, α -P³² dCTP, α -P³² dTTP, α -P³² dCTP, α -P³² dTTP, α -P³² dTTP, α -P³² dCTP, α -P³² dATP, α -P³² dTTP, α -P³² dCTP, α -P³² dTTP, α -P³² rCTP, α -P³² dCTP.

[00124] In Mg^{2+} buffer, substitution of a single rNTP with a dNTP reduces W.T. activity by 20 to >400-fold, but only modestly reduces the activity of Y639F. The rank order of the effect of a particular dNTP substitution on W.T. enzyme activity--
dGTP>dATP>dCTP>dTTP=dUTP--matches the order of their addition to the transcript. With the "2 dNTP" reactions the W.T. enzyme was most active when rCTP and rUTP were substituted

with dNTPs, corresponding to the 2 nucleotides added latest to the transcript. The W.T. enzyme was inactive with all other "2 dNTP" or "3 dNTP". In the "2 dNTP" reactions Y639F was least active in the dGTP, dATP reaction, corresponding to the 2 nucleotides incorporated first during transcription. In the "3 dNTP" reactions Y639F was least active in the dGTP, dATP, dCTP reaction, corresponding to the 3 nucleotides incorporated first during transcription.

[00125] In Mn^{2+} buffer both the W.T. enzyme and Y639F show a reduction in their sensitivity to substitution of dNTPs for rNTPs, consistent with an expectation of reduced substrate discrimination in Mn^{2+} buffer. There was, however, also a sharp reduction in overall activity with Mn^{2+} . The Mn^{2+} concentrations were varied over a wide range (from 20 mM to 150 μ M in 2-fold dilutions) to determine if an optimal Mn^{2+} concentration that would result in high activity could be identified, but similar activity was found at all Mn^{2+} concentrations tested (data not shown). Thus, while discrimination between rNTPs and dNTPs was less in Mn^{2+} buffer, Y639F is more active in Mg^{2+} buffer than in Mn^{2+} buffer with all NTP combinations examined. Similarly, the W.T. enzyme exhibits greatly reduced discrimination between rNTPs and dNTPs in Mn^{2+} buffer, but is modestly more active in Mn^{2+} buffer than in Mg^{2+} buffer only for certain combinations of dNTPs and rNTPs.

[00126] DNA and RNA synthesis on homopolymeric templates: T7 RNAP will synthesize poly(rG) RNAs on poly(dC) templates (Bonner, *et al.*, 1994; Ikeda and Richardson 1987). The activity of the W.T. and mutant polymerases were measured on poly(dI)·poly(dC) with rGTP, dGTP, dGTP+rGMP (see, Table III).

Table III: Relative activity on poly(dI)·poly(dC)

	W.T	Y639F	G640A	Y639A	Y639S
rGTP	1000	1000	240 (200-270)	145 (142-151)	48 (47-50)
dGTP	7.4 (5.4-12.5)	964 (684-1257)	<5	5.3 (4.8-5.5)	.6
dGTP+rGMP	25 (20-27)	1070 (816-1457)	<5	25 (17-30)	4.4

Numbers give mean and range from 3 experiments. Templates were at .2 mg/ml, polymerases at 10⁻⁸ M. Labeling NTPs were α -P³² rGTP, α -P³² dGTP, α -P³² rATP, α -P³² dATP, as appropriate. rNTPs or rNMPs at .5 mM; dNTPs at 1 mM.

[00127] The activity of T7 RNAP on poly(dI)·poly(dC) and poly(dC) is especially robust. Mutant polymerases that have greatly reduced activity on normal promoter templates still display

high activity on poly(dC) templates (Bonner, *et al.*, 1994). Therefore, two poorly active non-conservative tyrosine mutations were characterized on this template (Y639A and Y639S). Table III also presents results obtained with mutant G640A. Presentation of data for the latter mutant was selected because it is more comparable in activity to the Y639A/S mutants and because it is representative of a mutation which has marked effects on the kinetics of transcription but does not affect substrate discrimination, even though it is directly adjacent to Y639. It was found that all of the Y639 mutants exhibit reduced substrate discrimination as demonstrated by the fact that their differential activity in reactions containing dGTP or dGTP+rGMP vs. rGTP is less than for the W.T. enzyme or the G640A mutant. In fact Y639F displays similar activity with rGTP, dGTP, or dGTP+rGMP.

[00128] Since the nucleic acid synthesized by Y639F on poly(dI)·poly(dC) using dGTP is presumably composed solely of dNMPs it is expected to be resistant to alkaline hydrolysis (Schmidt and Tannhauser, 1945). Fig. 4 shows transcription by Y639F and W.T. polymerase with dGTP or rGTP on poly(dI)·poly(dC). Transcription reactions were carried out with poly(dI)·poly(dC) at .2 mg/ml and the indicated NTPs and polymerases. Reaction products were left untreated (-) or treated with 1 M NaOH for 5 hours at 37°C (+). Polymerase and NTP concentrations and electrophoresis as in Fig. 1. Labeling was with α -P³² rGTP (a, d, g, j) or α -P³² dGTP (other lanes). Fig. 4 shows the transcription products obtained with the W.T. or Y639F polymerases on poly(dI)·poly(dC) before and after treatment with alkali. With dGTP or dGTP+rGMP the W.T. enzyme is poorly active in the synthesis of long transcripts, however smears of heterogeneously sized, short transcripts can be observed in the reactions with the W.T. enzyme and the dGTP substrates (lanes e and f) while Y639F synthesizes higher levels of long transcripts which are retained near the top of these gels (lanes b and c). The presence of these short transcripts indicates that Y639F and the W.T. enzyme differ only in the degree to which they can utilize dNTPs. The W.T. can also initiate and extend transcripts with dGTP, but it is much less processive when using dNTPs than Y639F so its transcripts are much shorter. When these reactions are treated with base, degradation of the long transcripts made in the reactions with rGTP is observed and the amounts of short RNAs (presumably hydrolysis products) increase (lanes g and j). In the reactions in which dGTP or dGTP+rGMP were used as substrates no degradation of the transcripts by base treatment is observed, confirming that these transcripts are composed of dNMPs.

[00129] T7 RNAP as a reverse transcriptase or RNA replicase which initiates *de novo*: It has been reported that T7 RNAP can use both RNA and DNA templates (Konarska and Sharp,

1989). Therefore, it was determined if the Y639F mutant would use dNTPs when transcribing an RNA template (poly(rC), see Table IV).

Table IV: W.T. and Y639F activity on an RNA (poly(rC)) template

	.5 mM GTP	1 mM GTP	1 mM dGTP+.5 mM GMP
W.T.	1000	<.5	<.5
Y639F	505 (358-733)	62 (48-80)	116 (90-148)

Numbers give mean and range from 3 experiments. Template was at .2 mg/ml, polymerases at 10^{-6} M. Labeling NTPs were α -P³² rGTP, α -P³² dGTP.

[00130] Overall the activity of the W.T. and Y639F polymerases on poly(rC) with rGTP was 10-20-fold less than on poly(dC) (not shown), but this reduction did not preclude synthesis of high levels of RNA on poly(rC) by using higher polymerase concentrations than were used in the poly(dI)-poly(dC) reactions. When dGTP or dGTP+rGMP was used the W.T. enzyme was not measurably active on poly(rC), while the activity of Y639F was reduced by only ~4-fold (with dGTP+rGMP) or ~8-fold (with dGTP). Thus, both the W.T. and Y639F polymerases are capable of unprimed RNA-directed RNA polymerization while Y639F is also capable of unprimed reverse transcription.

[00131] DNA- and RNA-primed synthesis of DNA and RNA: In the assays described so far *de novo* initiated synthesis has been examined. T7 RNAP can also extend RNA primers. Therefore, the abilities of both polymerase to carry out DNA or RNA primed synthesis of DNA and RNA (Fig. 5) were examined.

[00132] Fig. 5 shows primed synthesis of DNA and RNA with Y639F and the W.T. polymerase. Transcription reactions contained end-labeled 12 base DNA or RNA primers of identical sequence (GGACACGGCGAA, SEQ ID NO: 4) hybridized to a DNA template (CCCGGGATGGAATGGAGTATTCGCCGTGTCCATGGCTGTAAGTATCC, SEQ ID NO: 5). Primer-template concentration was 10^{-5} M. Reactions contained the indicated polymerases (10^{-7} M) and NTPs. NTP concentrations and electrophoresis as in Fig. 1.

[00133] It was found that both the W.T. and Y639F polymerases can extend DNA and RNA primers with rNTPs, but extension of DNA primers was 2-3-fold less efficient than

extension of RNA primers. Y639F also extended DNA and RNA primers with dNTPs but ~4-fold less efficiently than with rNTPs.

[00134] The Y639F mutant does not exhibit greatly increased miscoding: examined were the miscoding properties of the W.T. and mutant T7 RNAPs by measuring the relative incorporation of labeled rGTP, rUTP, rATP, and rCTP on poly(dC) or poly(dT) templates in the presence of excess unlabeled rGTP or rATP, respectively (see, Table V). An increase in miscoding would be reflected in an increase in the rate of incorporation of the non-complementary NTP into RNA.

Table V: Relative rates of incorporation of complementary and non-complementary rNTPs on homopolymeric templates by W.T. and mutant polymerases

I. Template: poly(dC)					
	W.T.	Y639F	G640A	Y639A	Y639S
GTP/UTP	>2000 (Mg ²⁺ *) 53 (Mn ²⁺)*	>1760 55	>1320 n.d.	>184 n.d.	>50 n.d.
GTP/CTP	400 32	550 40	508 n.d.	>184 n.d.	>50 n.d.
GTP/ATP	233 9.3	338 8	388 n.d.	>184 n.d.	>50 n.d.
II. Template: poly(dT)					
	W.T.	Y639F	G640A	Y639A	Y639S
ATP/GTP	170 50	94 27	n.d. n.d.	n.d. n.d.	n.d. n.d.
ATP/UTP	121 21	94 20	n.d. n.d.	n.d. n.d.	n.d. n.d.
ATP/CTP	>340 77	>94 60	n.d. n.d.	n.d. n.d.	n.d. n.d.

Numbers are averages from 2 experiments and reflect the ratio of the percentages of labeled rNTPs incorporated into RNA in reactions in which unlabeled complementary rNTPs were in great excess (.5 mM) to labeled complementary or non-complementary rNTPs. Templates were at .1 mg/ml. Polymerases were at 10⁻⁸ M in Mg²⁺ buffer and 10⁻⁷ M in Mn²⁺ buffer.

*The upper number refers to results obtained in Mg²⁺ buffer, the lower number to results in Mn²⁺ buffer. n.d.: G640A, Y639A, Y639S were poorly active in Mn²⁺ buffer, or on poly(dT) under all conditions.

[00135] On poly(dC), the W.T., Y639F, and G640A polymerases incorporate rGTP into RNA at greater than 1300-2000-fold the rate of rUTP incorporation. Ribo-GTP is incorporated some 400-600-fold better than rCTP, and 200-400-fold better than rATP. Because of their lower activity it can be said that the relative rate of incorporation of rGTP on poly(dC) is 184-fold greater than the rate of incorporation of non-complementary rNTPs for Y639A, and 50-fold greater for Y639S. The use of Mn^{2+} instead of Mg^{2+} has been reported to increase miscoding for a number of polymerases (Tabor and Richardson, 1989; Nivogi and Feldman, 1981) so the effects of Mn^{2+} on miscoding by W.T. and Y639F polymerases were examined. In Mn^{2+} buffer the G640A, Y639A, and Y639S mutants were insufficiently active to allow accurate measures of miscoding. With the W.T. and Y639F polymerases the use of Mn^{2+} increases miscoding by 20-40-fold. However, the apparent rate of miscoding by Y639F remains similar to the w.t enzyme.

[00136] On poly(dT), high levels of activity allowing an accurate measure of miscoding frequencies could only be observed for Y639F and the W.T. polymerase in both Mg^{2+} and Mn^{2+} buffers. In Mg^{2+} buffer apparent miscoding rates on poly(dT) were higher than on poly(dC), but were similar for Y639F and the W.T. enzyme. In Mn^{2+} buffer miscoding rates were increased by ~5-fold, on average, but again rates were similar for W.T. and Y639F. However, on poly(dT) Y639F did show a reproducible ~2-fold increase, relative to the W.T. enzyme, in the ratio of the rates of rGTP to rATP incorporation.

[00137] Homopolymer assays have been used previously to measure miscoding by RNAPs (Nivogi and Feldman, 1981; Glazer, 1978; Blank, *et al.*, 1986) but it should be remarked that they can produce only upper bounds for miscoding frequencies. Measured miscoding rates could reflect contamination of the homopolymeric templates. It is also possible that the transcripts themselves could serve as templates and support incorporation of rNMPs non-complementary to the original template (i.e., the poly(rG) or poly(rA) transcripts made on poly(dC) or poly(dT) could subsequently support synthesis of poly(rC) or poly(rU) transcripts). Such caveats are less relevant to the miscoding observed in Mn^{2+} buffer since the change in divalent cation increases miscoding but cannot affect template composition. However, in Mg^{2+} buffer one should consider the measured miscoding frequencies to be upper bounds for the true rates of miscoding. Nevertheless, the results presented in Table V indicate that Y639F does not exhibit a gross increase in miscoding which would manifest itself as a clear increase in the incorporation of non-complementary rNMPs on homopolymeric templates.

[00138] The increased utilization of dNTPs by Y639F is due to both a decreased K_m and an increased k_{cat} for dNTPs: The K_m and k_{cat} of the W.T. and Y639F polymerases with rATP, rITP and 5 different dNTPs were measured (see, Table VI).

Table VI: Kinetic Constants for Y639F and the W.T. polymerase with rNTPs or dNTPs

		rATP	rITP	DTTP	dUTP	dCTP	dGTP	dATP
Y639F:	K_m	.053-.125	.034-.094	.038-.059	.052-.092	.92-1.6	.185-.264	.20-.35
	K_{cat}	150-210 s ⁻¹	180-200 s ⁻¹	70-110 s ⁻¹	70-130 s ⁻¹	50-90 s ⁻¹	30-50 s ⁻¹	50-70 s ⁻¹
W.T.:	K_m	.034-.068	.029-.059	.209-.262	4.4-9.0	4.3-13.5	.602-.701	2.0-5.0
	K_{cat}	190-220 s ⁻¹	170-230 s ⁻¹	26-29 s ⁻¹	25-39 s ⁻¹	9-14 s ⁻¹	5-9 s ⁻¹	6-9 s ⁻¹
Numbers give ranges from 3 experiments. The template was supercoiled PT75 AT 10 ⁻⁷ M, and polymerases were at 10 ⁻⁸ M.								

[00139] The K_m of the W.T. enzyme for dNTPs was much higher than previously reported values for the corresponding rNTPs and varied considerably for different dNTPs (Ikeda and Richardson, 1987; Patra, *et al.*, 1992). Notably the W.T. enzyme K_m values correlate with the rNTP/dNTP selectivity values presented in Table I. The selectivity of the W.T. enzyme with ribo- vs. deoxy-nucleotides was greatest for CTP, followed by ATP, and was the least for UTP, implying that an important component of the selectivity of the W.T. enzyme for rNTPs over dNTPs is a much higher K_m for dNTPs. For Y639F, the K_m values for these dNTPs are from ~3 to ~11-fold less, but the rank order of these K_m values (dCTP K_m > dATP K_m > dGTP K_m > dTTP K_m) is the same as for the W.T. enzyme. For Y639F, k_{cat} values for reactions with different dNTPs were only 2-4 fold less than for rNTPs, while the W.T. enzyme displayed k_{cat} values with dNTPs that were from ~6 to ~30-fold less than for rNTPs.

[00140] Elongation rates of Y639F in reactions containing a single dNTP: Elongation rates for Y639F in reactions with 4 rNTPs or 1 dNTP and 3 rNTPs were determined by analyzing aliquots, taken at 10 second intervals, from transcription reactions initiated on supercoiled PT75 by adding NTPs to otherwise complete reaction mixes (Fig. 6).

[00141] Fig. 6 shows relative elongation rates of Y639F in "4 rNTP" and "3 rNTP + 1 dNTP" reactions. The template was supercoiled pT75 at 5 x 10⁻⁷ M. Y639F polymerase was used at a concentration of 10⁻⁶ M. Reactions contained the indicated NTPs. Labeling was with α -P32-rATP (lane e) or α -P32-rGTP (other lanes). After initiation of the reactions aliquots were taken at 10 second intervals and analyzed on 1% agarose denaturing-formaldehyde gels. The figure shows the 20 second time point. The bars indicate the positions of λ DNA markers.

[00142] When analyzed on denaturing agarose gels (Fig. 6) the heterogeneously sized transcripts from these reactions are resolved as a smear with the trailing edge of the smear corresponding to transcripts initiated at t=0 and from which the maximal transcript elongation rate can be determined (Golomb and Chamberlin, 1974; Bonner, *et al.*, 1994). The following elongation rate reductions (relative to a '4 rNTP' reaction) were obtained for reactions containing

a 3 rNTPS and 1 dNTP:dATP, ~3-fold; dUTP, ~2-fold; dGTP ~1.5-fold; dCTP, 1-1.5-fold. Because of its poor activity in reactions with dNTPs it was difficult to determine the corresponding elongation rate reductions for the W.T. enzyme.

[00143] Other non-canonical nucleoside triphosphate substrates for mutant polymerases: Although wild-type T7 RNAP can not efficiently utilize dideoxy-NTPs as substrates, it was found that Y639 mutants of this enzyme can also use dideoxy-NTPs as substrates (see, Table VII).

Table VII: 2',3'-dideoxy NTP preferences of the Y639F mutant

ATP/ddATP	TTP/ddUTP	CTP/ddCTP	GTP/ddGTP
9.0	7.0	10.0	5.0

Numbers reflect the relative specificity (k_{cat}/K_m) for an NTP vs. the corresponding ddNTP. Relative specificities could not be evaluated for the wild-type T7 RNAP because the ddNTPs are such poor substrates, but these relative specificities appear to be at least 150-fold.

[00144] It was also found that Y639 mutants can use other non-canonical nucleoside triphosphates as substrates (see, Table VIII). Nucleic acids synthesized by incorporation of some non-canonical nucleotides, such as 2'-F-NTPs, may offer advantages in being more resistant to digestion by nucleases such as ribonucleases. Other uses and advantages of various non-canonical nucleotide substrates in methods of the present invention using the mutant polymerases will be apparent after examination of the specification, claims and drawings.

Table VIII: Activity of W.T. and Y639 mutants with NTPs containing different 2'-substituents

NTP	W.T.	Y639F	Y639M
UTP	100	95±6.7	50±1.2
2'-NH ₂ -UTP	5.9±.27	12±.41	3.6±.19
2'-F-UTP	3.1±.14	73±2.6	23±.72
2'-dUTP	2.4±.11	46±2.4	11±.46
CTP	100	103±2.3	54±3.9
2'-NH ₂ -CTP	34±.86	60±2.5	21±.38
2'-F-CTP	3.4±.22	63±3.1	47±.70
2'-dCTP	1.6±.16	57±1.7	32±1.1
ATP	100	96±3.0	51±1.3
2'-NH ₂ -ATP	18±.39	21±.75	.92±.035
2'-F-ATP	6.6±.12	50±1.4	9.7±.20
2'-dATP	2.7±.28	40±1.3	3.2±.11

Activity was determined with pT75 as template but with one of the rNTPs replaced with a dNTP or a 2'-modified NTP. The labeling NTP was UTP (in reactions with 2'-modified CTPs or ATPs) or CTP (in reactions with 2'-modified UTPs). Y639F and Y639M represent enzymes with substitutions of the wild-type (W.T.) tyrosine at position 639 by phenylalanine (F) or methionine (M), respectively.

[00145] Other mutations: It is conceivable that, in the absence of a bound rNTP, a hydrogen bond forms between Y639 and some other active site side chain. The possibility of an interaction between M635 and Y639 was tested since M635 and Y639 are close and M635 approaches the ribose in our models of NTP in T7 RNAP (Huang, *et al.*, submitted for publication). This position is methionine in the T7 RNAP class of RNAPs (McAllister, W.T., 1993), but is either tyrosine or phenylalanine in the homologous DNAPs, and in the DNAP mutants at this site (i.e., positions homologous to position 762 in *E. coli* DNAP I) that affect dNTP/ddNTP discrimination (Tabor, S., and Richardson, C.C., European patent application, 1994). While M635A, M635F or M635Y mutants had effects on NTP K_m , they did not affect 2'-group discrimination with respect to dNTPs and rNTPs in either wild-type or the Y639F T7 RNAP background. Additional studies will reveal whether these M635 mutations have other effects, such as effects on discrimination at the 3'-position of the sugar, or effects on discrimination with respect to NTPs with other substituents, such as fluorine at the 2' position of the sugar. Also, studies on similar double mutations at the homologous sites in the homologous class I DNAPs will reveal the effects of such double mutations on discrimination at the 2'- and 3'-positions of the sugar; specifically, these studies will reveal whether class I DNAPs having a phenylalanine at the position homologous to amino acid position 766 in *E. coli* DNAP I have

reduced discrimination for rNTPs if the amino acid is methionine or tyrosine at the position homologous to amino acid position 762 in *E. coli* DNAP I.

[00146] Provided that class I DNAP mutants which have a reduced discrimination for rNTPs compared to dNTPs can be obtained, it will be possible to use such DNAP mutants to carry out the methods of the present invention that comprise nucleic acid synthesis from a nucleic acid primer, at least part of which is sufficiently complementary to a template nucleic acid to hybridize therewith and to be extended by the polymerase. An especially preferable use for such mutant DNA polymerases would be to carry out Partial Ribo-substitution sequencing reactions from primers, whether labeled by any of the methods known in the art, or unlabelled. Since the sequence-delimiting rNTP nucleotides for the Partial Ribo-substitution Reaction do not terminate the growing phosphodiester chain when they are incorporated during nucleic acid synthesis, the DNA synthesis for Partial Ribo-substitution can occur simultaneous with and be identical to nucleic acid synthesis for another procedure, such as NASBA, 3SR, TMA or another similar method, provided that the mutant DNAP with reduced discrimination for rNTP compared to dNTPs is thermostable, or the PCR strand displacement amplification or other methods for nucleic acid amplification involving nucleic acid synthesis from a primer. The nucleic acid products containing the sequence-delimiting rNMPs can then be cleaved by treatment with a chemical base or a ribonuclease, and analyzed by methods known in the art to obtain the nucleotide-specific pattern of bands or, provided that a Partial Ribo-substitution Reaction is carried out for each of the four nucleotides, the complete sequence of the nucleic acid. Also, provided that primers with distinguishable non-radioactive labels are used, multiple Partial Ribo-substitution sequencing reactions may be carried out simultaneously in the same reaction mixture and run and read in the same lane of a polyacrylamide gel or capillary tube or other matrix for separating the fragments based on size, as is the case for all of the methods of the present invention described herein.

C. Discussion

[00147] Our results reveal that mutations of tyrosine 639 in T7 RNAP reduce the ability of the polymerase to discriminate between rNTPs and dNTPs. A conservative mutation which removes the tyrosine hydroxyl but retains the phenolic ring (Y639F) exhibits W.T. activity but an average reduction of ~20-fold in the selectivity for dNTPs over rNTPs (Table I). Non-conservative mutations of this tyrosine (Y639A/S) also display decreased rNTP/dNTP discrimination (Table III), but are less active than the W.T. enzyme. Replacement of an rNTP by a dNTP typically reduces Y639F transcript elongation rates by only a factor of two. Tyrosine

639 is conserved in a large number of DNA-directed RNA and DNA polymerases (Delarue, et al., 1990). In DNAP I, mutations of the Y766 to serine and phenylalanine have been characterized (Polesky, et al., 1989; Carrol, et al., 1991). The Y766S mutation was alone amongst a number of active site mutations characterized in decreasing DNAP I fidelity (increasing miscoding). The Y766F mutation displayed W.T. fidelity and activity. Similarly, the T7 RNAP Y639F mutant displays W.T. kinetics (Bonner, et al., 1992, 1994; Woody, et al., 1994), and the only effect which can be identified for this mutation in T7 RNAP is the reduced substrate discrimination reported here. Thus, while T7 RNAP Y639F showed decreased dNTP/rNTP selectivity, it did not exhibit increased miscoding as assessed by incorporation of non-complementary NTPs on homopolymeric templates.

[00148] The Y639 T7 RNAP mutations present us with, in one sense, the functional unification of polymerases to go along with their structural unification. The active site of W.T. T7 RNAP is forgiving with regards to template structure (RNA or DNA) or mode of initiation (primed or *de novo*). A mutation which relaxes the substrate selectivity of this polymerase further expands the range of activities which it can display *in vitro*. Depending on the substrates and templates presented to it, the Y639F T7 RNAP can act as an RNA- or DNA-directed RNA or DNA polymerase in primed or *de novo* initiated reactions. Thus it can display a variety of activities normally associated with distinct polymerases, including some entirely novel activities such as *de novo* initiated reverse transcription or mixed dNMP/rNMP polymer synthesis.

Example 2: A mutant SP6 RNA Polymerase as a DNA Polymerase

[00149] After the observations made above with T7 RNAP, bacteriophage SP6 RNA polymerase was examined to determine whether the DNA synthesis properties observed for the mutant T7 RNAP could, as expected, be extended to other mutant polymerases. Bacteriophage SP6 is a lytic phage which infects the bacterial species *Salmonella typhimurium* (Butler and Chamberlin, 1982). SP6 phage resembles *E. coli* phage T7 and their genomes are comparable in size, gene organization and pattern of gene expression (Kassavetis, et al., 1982).

[00150] The phage encoded RNA polymerases are very similar in size (Butler and Chamberlin, 1982) and amino acid sequence (Katani, et al., 1987).

[00151] The homologous tyrosine at position 639 in T7 RNA polymerase is readily identified at position 631 in SP6 RNA polymerase (Fig. 7). Substitution of tyrosine 631 with phenylalanine in the SP6 RNA polymerase was expected to confer the same phenotypic changes in catalytic properties in this enzyme as were demonstrated for Y639FT7 RNA polymerase (Example 1).

[00152] Localized Mutagenesis. Refer to Fig. 7 for a summary of the amino acid and nucleotide sequence surrounding TYR631 in the SP6 RNA polymerase gene. Mutagenesis of the Y631 residue may be accomplished by the method of Kunkel, et al. (1991). Alternatively, one of the many other methods for mutagenesis known to those of skill in the art may be used. The amino acid and nucleotide sequences of the resulting TYR631PHE mutant SP6 RNA polymerase are also given in Fig. 7. As shown, the A residue at position 2 in codon 631 of the SP6 RNA polymerase gene was changed to a T. This results in the loss of the single NdeI restriction enzyme site which is present in the wild-type gene, permitting identification of mutant clones.

Preparation and Purification of Mutant RNA Polymerase

[00153] A single clone was selected in which the NdeI site was missing and which expressed SP6 RNA polymerase activity. Several liters of LB + Amp were grown from a pUC18 Y631F SP6 RNA polymerase clone overnight at 37°C. Cells were harvested, lysed and Y631F mutant SP6 RNA polymerase was purified approximately according to standard methods (Butler and Chamberlin, 1983).

Transcription Reactions

[00154] To verify that Y631F SP6 RNAP had the desired phenotype, *in vitro* transcription reactions were done where one of the four rNTPs was substituted by the corresponding dNTP (Sousa and Padilla, 1995). As expected, the Y631F SP6 RNAP mutant displayed reduced dNTP/rNTP discrimination compared with wild-type SP6 RNAP, similar to that observed for the Y639 mutant of T7 RNAP.

[00155] In a standard *in vitro* transcription reaction using the four ribonucleoside triphosphates (rATP, rGTP, rCTP, and rUTP), both enzymes, the wild-type SP6 RNA polymerase as well as the Y631F mutant, synthesized the correct 1.4 kb transcript and in the expected amounts as visualized on gels. However, if one of the four ribonucleoside triphosphates, rGTP for example, is completely substituted by dGTP and *in vitro* transcription reactions are done with the wild-type and mutant enzymes, no transcript is made by the wild-type enzymes. However, the mutant enzyme makes the expected full length transcript in good yield as observed on agarose gels.

Example 3: Activity and Uses of Phage RNA Polymerases With Second Mutations Comprising an Alternative Amino Acid In Place of the Histidine at the Position Corresponding to H784 of T7 RNAP for Synthesis of RNA Containing Non-Canonical Nucleotides

Background

[00156] RNAPs that incorporate non-canonical nucleotides are useful for making RNA molecules for a variety of applications, including, but not limited to, applications previously described herein. In one preferred embodiment, mutant RNAPs are used for synthesis of RNA molecules containing non-canonical nucleotides for use as probes for nucleic acid arrays or for using in silencing genes *in vivo* by means of acting as an interference RNA or RNAi, including, but not limited to, a small interference RNA or siRNA. In some embodiments, the synthesized nucleic acids with non-canonical nucleotides are resistant to non-specific nucleases or have other special properties.

[00157] In another preferred embodiment, mutant T7-like RNAPs can be used to synthesize RNA containing non-canonical nucleotides for use in a method referred to “immuno-detection amplified by T7 RNA polymerase” or “IDAT” (Zhang, H-T, *et al.*, “Protein quantification from complex protein mixtures using a proteomics methodology with single-cell resolution.” *Proc. Natl. Acad. Sci. USA* 98:5497-5502, 2001). In still another preferred embodiment, mutant T7-like RNAPs are used to synthesize RNA containing non-canonical nucleotides in order to analyze specific binding of the nucleic acid to cellular proteins, such as receptors, drug target candidates, drug candidates, or other biomolecules for use in proteomic, pharmacogenomic, or drug screening research. In certain aspects of proteomics research it is particularly desirable to use a mutant RNAP that can incorporate non-canonical nucleotides which are resistant to certain nucleases and/or which have properties such as having a chemical moiety which can be more easily linked to proteins or other bio-molecules, drug target candidates, or drug candidates which are in close proximity in the cell or when non-covalently bound to the nucleic acid, directly or indirectly, on a nucleic acid or protein array. By way of example, but not of limitation, an azido group in the 2'-position of a non-canonical nucleotide provides methods for linkage of a nucleic acid which contains it to other nucleic acids, proteins, bio-molecules, drug target candidates, or drug candidates in close proximity. For example, such a group can be used to link the nucleic acid to other bio-molecules by chemical or photochemical means.

[00158] As disclosed previously herein, when the tyrosine in the conserved motif with the consensus sequence ...K-----YG... of T7-like RNAPs is mutated to phenylalanine, the mutant

RNAP, including, but not limited to the Y639F mutant of T7 RNAP, the Y631F mutant of SP6 RNAP, or the Y573F mutant of T3 RNAP, readily utilizes 2'-deoxy, 2'-NH₂, and 2'-F NTPs as substrates. However, these mutant RNAPs do not readily utilize non-canonical nucleoside triphosphates with bulkier 2'-substituents or facilitate incorporation of nucleoside triphosphates with modifications at other positions.

[00159] This Example discloses that introduction of a second mutation in a T7-like RNAP which alters the conserved histidine corresponding to H784 of T7 RNA polymerase, H779 of SP6 RNAP, or H785 of T3 RNAP, or the terminal histidine in a conserved amino acid sequence, which conserved sequence comprises at least 2, 3, 4, 5 or 6 of the amino acids in the sequence "APNFEVH," of these or other T7-like RNAPs, markedly enhances utilization of non-canonical NTPs with bulky 2'-substituents, such as, but not limited to, 2'-OMe and 2'-N₃ (i.e. "azido") groups.

[00160] In preferred embodiments of the invention, the corresponding histidine is mutated to alanine, but the invention also includes embodiments in which the histidine is mutated to another amino acid that occupies less space than the side-chain of histidine and which permits synthesis of RNA which contains non-canonical nucleotides with bulky 2'-substituents, such as, but not limited to, 2'-OMe and 2'-azido groups. Other preferred embodiments of the invention use mutant RNAP enzymes with a second mutation which converts the histidine corresponding to H784 of T7 RNAP or the like histidine of another T7-like RNAP, as described above, so as to also permit improved synthesis of nucleic acids using non-canonical nucleotides, such as, but not limited to, dNTPs or other non-canonical nucleotides which also have modifications at other positions on the nucleic acid base or the sugar moiety and/or NTPs with modifications on the nucleic acid base or the sugar moiety at positions other than the 2'-position. By way of example, but not of limitation, the modifications on the nucleic acid base or sugar moiety of the nucleotides used in these embodiments can be a fluorescent moiety, such as, but not limited to, a Cy5 dye which is covalently linked to a cytosine or uridine base via an aminoalkyl linkage, including, for example, FluoroLink™ Cy5-dCTP or FluoroLink™ Cy5-UTP (Amersham Biosciences).

[00161] Without limiting the invention, it is believed that the Y639F mutation of T7 RNAP does not overcome the barrier to synthesis of nucleic acids with bulkier substituents at the 2'-position of non-canonical nucleotides because the tyrosine of Y639 is positioned so as to discriminate the structure of the substrate NTP, but is not involved in sensing the structure of the 3'-NMP of the RNA. The crystal structure of a T7 RNAP transcription complex (Fig. 1) reveals that the side-chain of histidine 784 occupies the minor groove side of the 3'-rNMP:template base-

pair (Cheetham, G.M.T., and Steitz, T.A., "Structure of a transcribing T7 RNA polymerase initiation complex," Science 286:2305-2308, 1999). Though the H784 histidine is well-conserved in T7-like phage RNAPs, it was previously found that an H784A mutant had near wild-type activity, but exhibited enhanced extension of RNAs with mispaired 3'-termini (Huang, J., et al., "Misincorporation by wild-type and mutant T7 RNAPs," Biochemistry 39:11571-11580, 2000).

[00162] The present invention discloses methods for using RNAPs having a second mutation, which mutation comprises or corresponds to the H784A mutation of T7 RNAP, to synthesize nucleic acid transcripts containing bulkier non-canonical NMPs. Without limiting the invention, it appears that replacement of histidine, with its bulky side-chain, by alanine makes room for non-canonical nucleotides with bulkier minor groove substituents.

Materials and Methods

[00163] RNAPs were prepared as described (Briebe, LG, and Sousa, R. Biochemistry, 39: 919-923, 2000). Transcription reactions were carried out for 30 min. at room temperature in 40 mM Tris-Cl, pH 8.0, 6 mM MgCl₂, 10 mM NaCl, 2 mM spermidine, with templates at 10⁻⁷ M and RNAPs at 2x10⁻⁷ M. All NTPs were used at 0.5 mM, and transcripts were labeled by inclusion of 1% v/v 800 Ci/mM μ M³²P-GTP in the reaction. Modified NTPs were from Trilink Biotechnologies. Reactions were stopped by addition of an equal volume of 90% formamide, 50 mM EDTA, 0.01% xylene cyanol, resolved by electrophoresis in 20% acrylamide, 1% bis-acrylamide, 1X TBE gels, and visualized on a Molecular Dynamics Phosphorimager.

[00164] In separate experiments to evaluate the ability of the Y639F/H784A double mutant versus the Y639F single mutant of T7 RNAP to incorporate nucleoside triphosphates with substituents at other positions than the 2'-position of the sugar, *in vitro* transcription reactions are carried out with each mutant enzyme using a T7 Control Template from an AmpliScribe™ T7 High Yield Transcription Kit (EPICENTRE) and either FluoroLink™ Cy5-UTP or FluoroLink™ Cy5-dCTP (both from Amersham Biosciences) in the reaction mixture in addition to the corresponding canonical nucleotide, UTP or CTP. Following transcription, the product is precipitated with ammonium acetate and the relative amount of Cy5 dye in the transcripts are determined based on the A650/A260 absorption ratios.

Results

[00165] The ability of the wild-type, Y639F and Y639F/H784A mutants to utilize 2'-OMe-UTP and 2'-OMe-CTP, and 2'-azido-UTP and 2'-azido-CTP in RNA synthesis was tested. On a

Bgl II-cut pPK5 (Mentesanas, P.E., *et al.*, "Characterization of Bacteriophage T7 RNA Polymerase Elongation Complexes," *J. Mol. Biol.* 302:1049-1062, 2000) template in which CTP and UTP are first incorporated at +14 and +15, respectively, the wild-type enzyme can synthesize 34 nucleotide (nt) run-off transcripts in reactions in which UTP is replaced with 2'-OMe-UTP (Fig. 2a, lane 2) or in reactions in which CTP is replaced with 2'-OMe-CTP (Fig. 2a, lane 3). Relative to the reaction with 4 NTPs (Fig. 2a, lane 1), yields of run-off transcripts are reduced 6-fold and 21-fold in the 2'-OMe-UTP and 2'-OMe-CTP reactions, respectively. When both UTP and CTP are replaced with the corresponding 2'-OMe-NTPs, run-off transcription with the wild-type enzyme is undetectable (Fig. 2a, lane 4). Terminated transcripts are seen in these reactions at positions where 2'-OMe-UMP or 2'-OMe-CMP are incorporated (Fig. 2a, lanes 2-4). The Y639F mutation enhances utilization of the 2'-OMe-NTPs: relative to reactions with 4 NTPs (Fig. 2a, lane 5) runoff transcription with Y639F is reduced, respectively: 2-fold, 8-fold, and 55-fold in reactions with 2'-OMe-UTP, 2'-OMe-CTP, or both (Fig. 2a, lanes 6-8). With the Y639F/H784A double mutant the reduction in run-off transcript synthesis is <15% in reactions with 2'-OMe-UTP (Fig. 2a, lane 10), and only 3-fold and 6-fold in reactions with, respectively, 2'-OMe-CTP or both 2'-OMe-UTP and 2'-OMe-CTP (Fig. 2a, lanes 11, 12).

[00166] Identical experiments were carried out with 2'-azido-UTP or 2'-azido-CTP replacing UTP or CTP, respectively (Fig. 2b). With the wild-type enzyme, run-off transcription was reduced 15- and 50- fold in reactions with 2'-azido-UTP or 2'-azido-CTP, respectively (Fig. 2b; lanes 2, 3), and was undetectable in reactions with both 2'-azido-UTP and 2'-azido-CTP (Fig. 2b; lane 4). With Y639F, use of 2'-azido-UTP or 2'-azido-CTP reduced run-off transcription by 70% (Fig. 2b; lanes 6, 7) compared to reactions with 4 NTPs (Fig. 2b; lane 5), while use of both 2'-azido-UTP and 2'-azido-CTP reduced runoff transcription by 11-fold (Fig. 2b, lane 8). With Y639F/H784A use of a single azido-modified NTP reduced run-off transcription by only 40% (Fig. 2b; lanes 10, 11), while use of two 2'-azido-NTPs reduced it by 3-fold (Fig. 2b; lane 12).

[00167] With the pPK5 template, UMP and CMP are not incorporated into the first 9 nt of RNA. Incorporation of non-canonical NMPs during synthesis of the first 9 nt of RNA is expected to cause greater reductions in run-off transcript synthesis because the RNA is readily released from the transcription complex during this initial phase of transcription. Therefore, if incorporation of a non-canonical NMP is slow, the RNA will usually dissociate before the NMP is incorporated. Once the RNA is greater than 9 nt in length the transcription complex becomes much more stable, and is unlikely to release the RNA during the time required for incorporation of a non-canonical NMP. To see whether NMPs with bulky 2'-substituents could be incorporated during the initial phase of transcription a Hind-III linearized pT75 (Tabor, S. and Richardson, C.

C., “A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes,” Proc. Natl. Acad. Sci. USA 82:1074-1078, 1985) was used. On this template, UMP is first incorporated at +14 while CMP is incorporated at +6 and +7. In reactions with 2'-OMe-UTP run-off transcript synthesis is reduced 11-fold, 5-fold, and <2-fold with the wild-type, and the Y639F, and Y639F/H784A mutants, respectively (Fig. 2c; lanes 2, 6, 10). However, in reactions with 2'-OMe-CTP run-off transcription is undetectable with the wild-type or Y639F enzymes (Fig. 2c; lanes 3, 4, 7, 8), and is reduced ~700-fold with Y639F/H784A (Fig. 2c; lanes 11, 12).

[00168] Run-off transcript yields in reactions with *Hind*III-cut pT75 and 2'-azido-UTP are reduced 18-, 8.5-, and <2-fold with wild-type and Y639F and Y639F/H784 mutants, respectively (Fig. 2d; lanes 2, 6, 10), relative to reactions with 4 NTPs (Fig. 2d; lanes 1, 5, 9). In reactions with 2'-azido-CTP, or both 2'-azido-CTP and 2'-azido-UTP, run-off transcription is undetectable with the wild-type enzyme (Fig. 2d; lanes 3, 4), is reduced 65-fold with Y639F (Fig. 2d; lanes 7, 8), but is reduced only 7-fold with Y639F/H784A (Fig. 2d; lanes 11, 12).

[00169] In the experiments to evaluate the ability of the Y639F/H784A double mutant versus the Y639F single mutant of T7 RNAP to incorporate nucleoside triphosphates with substituents at other positions than the 2'-position of the sugar, the A650/A260 absorption ratios of the *in vitro* transcription products are found to be significantly higher using the Y639F/H784A double mutant enzyme compared to the Y639F single mutant of T7 RNAP when either FluoroLink™ Cy5-UTP or FluoroLink™ Cy5-dCTP is used in the transcription reaction in addition to the corresponding canonical nucleotide, UTP or CTP, indicating that the double mutant is better able to synthesize transcripts containing the Cy5 moiety than the single mutant.

Discussion

[00170] By introducing an H784A substitution into a Y639F background, a polymerase was obtained with an enhanced ability to incorporate NMPs with bulky 2'-substituents into RNA. In reactions with 2'-OMe- or 2'-azido- modified NTPs, yields of run-off transcripts, relative to reactions with 4 canonical NTPs, are markedly increased with the double mutant and premature termination products are greatly reduced or eliminated. Incorporation of the modified NMP is more efficient if it occurs during elongation, rather than during the poorly processive initiation phase of transcription when the RNA is less than 9 nt in length. Given the position of the H784 side-chain in the crystal structure of an initial transcript complex (Cheetham, G.M.T. and Steitz, T.A., Science 286:2305-2308, 1999), it is believed likely, without limiting the invention, that the effect of the H784A mutation is due primarily to enhanced extension of RNAs containing a

modified NMP at the 3'-terminus. This would be consistent with the observation that the H784A mutation enhances extension of RNAs with mispaired 3'-termini, but does not increase incorporation of mispaired NMPs (Huang, J., et al., Biochemistry 39:11571-11580, 2000).

[00171] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

[00172] The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference in their entirety.

REFERENCES

U.S. PATENTS

U.S. patent 4,683,202

U.S. patent 4,965,188

INTERNATIONAL PATENTS

Tabor, S and Richardson, C.C. (Filed 24.11.94) European Patent Application Number 94203433.1; Publication Number 0 655 506 A1.

PUBLICATIONS

Astatke, M., Grindley, N.D.F., and Joyce, C.M. (1995) J. Biol. Chem. 270:1945-1954.

Axelrod, V.D, Vartikyan, R.M., Aivazashvilli, V.A., and Bebelashvilly, R.S. (1978) Nucleic Acids Res. 5:3549-3563.

Axelrod, V.D, and Kramer, P.R. (1985) Biochemistry 24:5716-5723.

Barnes, W.M. (1978) J. Mol. Biol. 119:83-99.

Beese, L.S., Friedman, J.M., and Steitz, T.A. (1993) Biochemistry 31:9636.

Blank, A., Gallant, J.A., Burgess, R.R., Loeb, L.A. (1986) Biochemistry 25:5920-5928.

Bonner, G., Patra, D., Lafer, E.M., and Sousa, R. (1992) EMBO J. 11:3767-3775.

Bonner, G., Lafer, E.M., and Sousa, R. (1994) J. Biol. Chem. 42:25120-25128.

Briebe, L.G., et al., (2002) Biochemistry 41:5144-5149.

Butler, E.T. and Chamberlin, M.J. (1982) J. Biol. Chem. 257:5772-5778.

Carroll, S.S., Cowart, M., and Benkovic, S.J. (1991) Biochemistry 30:804-13.

Cazenave, C., and Uhlenbeck, O.C. (1994) Proc. Natl. Acad. Sci. USA 91:6972-6976.

Chapman, K.A., and Burgess, R.R. (1987) Nucl. Acids Res. 15:5413-5426.

Chapman, K.A., Gunderson, S.I., Arnello, M., Wells, R.D., and Burgess, R.R. (1988) Nucl. Acids Res. 16:4511-4530.

Compton, J. (1991) Nature 350:91-92.

Cotton (1993) Mutation Res. 285:125-144.

Cunningham, P.R., and Ofengand, J. (1990) BioTechniques 9(6):713-714.

Duck, P.G., Alvarado-Urbina, B., Burdick, B., and Collier, B. (1990) BioTechniques 9:142-148.

Fahy, et al. (1991) PCR Methods & Applications 1:25-33.

Glazer, R.I. (1978) Nucleic Acids Res. 5:2607-2616.

Golomb and Chamberlin, (1974) J. Biol. Chem. 249:2858-2863.

- DeLarue, M., Poch, O., Tordo, N., Moras, D., and Argos, P. (1990) J. Protein Engineering 10:461-467.
- Hill, C.S. (1996) "Gen-Probe Transcription-mediated Amplification System Principles," Tech. Bulletin No: L137/01/96, of Gen-Probe, Inc.
- Huang, Y., et al. (1997) Biochemistry 36:8231-8242.
- Ikeda, R.A., Richardson, C.C. (1987) J. Biol. Chem. 262:2800-3808.
- Jacobo-Molina, A, Ding, J., Nanni, R.G., Clark, A.D., Lu, X., Tantillo, C., Williams, R.L., Kramer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S.H. and Arnold, E. (1993) Proc. Natl. Acad. Sci. (USA) 90:6320.
- Kassavetis, G.A., Butler, E.T., Roulland, D., and Chamberlin, M.J. (1982) J. Biol. Chem. 257:5779-5788.
- Katani, H., Ishizaki, Y., Hiraoka, N., and Obayashi, A. (1987) Nucleic Acids Res. 15:2653-2664.
- Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., and Steitz, T. A. (1992) Science 256:1783-1790.
- Konarska, M.M., and Sharp, P.A. (1989) Cell 57:423-431.
- Kramer, F.R., and Mills, D.R., (1978) Proc. Natl. Acad. Sci. (USA) 75:5334-5338.
- Kuchta, R.D., Mizrahi, V., Benkovic, P.A., Johnson, K.A., and Benkovic, S.J. (1987) Biochemistry 26:8410-8417.
- Kunkel, T.A., Bebenek, K., and McClary, J. (1991) Methods in Enzymology 204:125.
- Makarova, O.V., Makarov, E.M., Sousa R., and Dreyfus, M. (1995) Proc. Natl. Acad. Sci. USA. (submitted).
- Martin, C.T., and Coleman, J.E. (1987) Biochemistry 26:2690-2696.
- Martin, C.T., Muller, D.K., and Coleman, J.E. (1988) Biochemistry 27:3966-3974.
- Martin, C.T., and Coleman, J.E. (1989) Biochemistry 28:2760-2762.
- McAllister, W.T. (1993) Cellular & Molecular Biol. Res. 39:385.
- McClure, W.R., and Chow, Y. (1980) Methods of Enzymology 64:277-297.
- Milligan, J.F., Groebe, D.R., Witherwell, G.W., and Uhlenbeck, O.C. (1987) Nucl. Acids. Res. 15:8783-8798.
- Mizrahi, V., Henrie, R.N., Marlier, J.F., Johnson, K.A., and Benkovic, S.J. (1985) Biochemistry 24:4010-4018.
- Moroney, S.E., and Picirilli, J.A. (1991) Biochemistry 30:10343-10349.
- Niyogi, S.K., and Feldman, R.P. (1981) Nucleic Acids Res. 9:2615-2627.
- Myers and Gelfand, D. (1991) Biochemistry 30:7661-7666.

- Padilla, R. and Sousa, R. (1999) Nucl. Acids Res. 27:1561-1563.
- Padilla, R. and Sousa, R. (2002) Nucl. Acids Res. 30: No. 24, e138.
- Patra, D., Sousa, R., and Lafer, E.M. (1992) J. Mol. Biol. 224:307-318.
- Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H., and Kraut, J. (1994) Science 264:1891.
- Polesky, A.H., Steitz, T.A., Grindley, N.D., and Joyce, C.M. (1989) J. Biol. Chem. 265:14579-14591.
- Ricchetti, M. and Buc, H. (1993) EMBO J. 12:387-396.
- Ryder, S.P., et al. (2000) Methods Enzymol. 317:92-109.
- Sambrook, J., Fritsch, E.F., and Maniatis, T., (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Sanger, et al. (1977) Proc. Natl. Acad. Sci. (USA) 74:5463-5468.
- Sawaya, M.R., Pelletier, H., Kumar, A., Wilson, S.H., and Kraut, J. (1994) Science 264:1930.
- Schmidt, G. and Tannhauser, S.J. (1945) J. Biol. Chem. 159:83-89.
- Shi, Y., Gamper, H., and Hearst, J.E. (1988) J. Biol. Chem. 263:527-534.
- Sousa, R., and Padilla, R. (1995) EMBO J. 14:4609-4621. (Incorporated by reference as if set forth herein.)
- Sousa, R., Lafer, E.M., and Wang, B.-C. (1991) J. Crystal Growth 110:237-246.
- Sousa, R., Chung, Y.J., Rose, J.R., and Wang, B.C. (1993) Nature 364:593-599.
- Steitz, T.A., Smerdon, S.J., Jager, J., and Joyce, C.M. (1994) Science 266:2022-2025.
- Tabor, S., and Richardson, C.C. (1990) J. Biol. Chem. 265:8322.
- Tantillo, C., Jianoing, D., Jacobo-Molina, A., Nanni, R.G., Boyer, P.L., Hughes, S.H., Pauwels, R., Andiries, K., Janssen, P.A.J., and Arnold, E. (1994) J. Mol. Biol. 243:369-387.
- Tabor, S., and Richardson, C.C. (1989) Proc. Natl. Acad. Sci. USA 82:1074-1078.
- Tabor, S., and Richardson, C.C. (1985) Proc. Natl. Acad. Sci. USA 82:1074-1078.
- Osumi-Davis, P.A., Sreerama N., Volkin, D.B., Middaugh, C.R., Woody, R.W., and Woody, A.Y. (1994) J. Mol. Biol. 237:5-19.